

Citation for published version:

Zammit-Mangion, A, Rougier, J, Schön, N, Lindgren, F & Bamber, J 2015, 'Multivariate spatio-temporal modelling for assessing Antarctica's present-day contribution to sea-level rise', *Environmetrics*, vol. 26, no. 3, pp. 159-177. <https://doi.org/10.1002/env.2323>

DOI:

[10.1002/env.2323](https://doi.org/10.1002/env.2323)

Publication date:

2015

Document Version

Publisher's PDF, also known as Version of record

[Link to publication](#)

Publisher Rights

CC BY

University of Bath

Alternative formats

If you require this document in an alternative format, please contact:
openaccess@bath.ac.uk

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Multivariate spatio-temporal modelling for assessing Antarctica's present-day contribution to sea-level rise

Andrew Zammit-Mangion^{a,b,*}, Jonathan Rougier^b, Nana Schön^a, Finn Lindgren^c and Jonathan Bamber^a

Antarctica is the world's largest fresh-water reservoir, with the potential to raise sea levels by about 60 m. An ice sheet contributes to sea-level rise (SLR) when its rate of ice discharge and/or surface melting exceeds accumulation through snowfall. Constraining the contribution of the ice sheets to present-day SLR is vital both for coastal development and planning, and climate projections. Information on various ice sheet processes is available from several remote sensing data sets, as well as *in situ* data such as global positioning system data. These data have differing coverage, spatial support, temporal sampling and sensing characteristics, and thus, it is advantageous to combine them all in a single framework for estimation of the SLR contribution and the assessment of processes controlling mass exchange with the ocean.

In this paper, we predict the rate of height change due to salient geophysical processes in Antarctica and use these to provide estimates of SLR contribution with associated uncertainties. We employ a multivariate spatio-temporal model, approximated as a Gaussian Markov random field, to take advantage of differing spatio-temporal properties of the processes to separate the causes of the observed change. The process parameters are estimated from geophysical models, while the remaining parameters are estimated using a Markov chain Monte Carlo scheme, designed to operate in a high-performance computing environment across multiple nodes. We validate our methods against a separate data set and compare the results to those from studies that invariably employ numerical model outputs directly. We conclude that it is possible, and insightful, to assess Antarctica's contribution without explicit use of numerical models. Further, the results obtained here can be used to test the geophysical numerical models for which *in situ* data are hard to obtain. © 2015 The Authors. *Environmetrics* published by John Wiley & Sons Ltd.

Keywords: Multivariate modelling; parallel MCMC; sea-level rise; spatio-temporal statistics; stochastic partial differential equations

1. INTRODUCTION

Antarctica is the largest fresh-water resource on Earth and, potentially, the largest contributor to global sea-level rise (SLR). Its ice sheet is widely believed to be losing on the order of 50 Gt year⁻¹ (Shepherd *et al.*, 2012; Gunter *et al.*, 2014; Rignot *et al.*, 2008), and this rate is likely to increase (Rignot *et al.*, 2011b). SLR has global implications and accurate assessment of present-day rates, sources and their uncertainties are essential for prediction, and hence the development and implementation of effective urban, eco-preservation and flood defence strategies (Vaughan *et al.*, 2013).

A review of the difficulties associated with estimating the SLR contribution from Antarctica and the assumptions of current approaches that tackle this problem is given in Zammit-Mangion *et al.* (2014). The primary difficulty is that of under-determination (lack of mixture diversity), because the number of observed linear combinations (using satellite and global positioning system (GPS) instruments) is less than the number of processes influencing the instruments. Zammit-Mangion *et al.* (2014) showed that under-determination, for this problem, can be tackled using a hierarchical modelling framework (Cressie and Wikle, 2011). Their approach, which as a proof of concept focused on only a part of the Antarctic ice sheet, took advantage of differing spectral characteristics to improve the separation of the individual processes

* Correspondence to: Andrew Zammit-Mangion, School of Geographical Sciences, University of Bristol, BS8 1SS, U.K. E-mail: A.ZammitMangion@bristol.ac.uk

^a School of Geographical Sciences, University of Bristol, Bristol, BS8 1SS, U.K.

^b Department of Mathematics, University of Bristol, Bristol, BS8 1TW, U.K.

^c Department of Mathematical Sciences, University of Bath, Bath, BA2 7AQ, U.K.

This is an open access article under the terms of the Creative Commons Attribution 4.0 License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

from the data. The analysis was time invariant and considered the mean rate of height change of each process, between 2003 and 2009, at every spatial location, to be the variable of interest.

In this article, we extend Zammit-Mangion *et al.* (2014) in several ways. (i) We consider *spatio-temporal* changes in height change instead of just spatial rates. This has two benefits, first differing temporal characteristics across the processes play a key role in improving the predictive performance on the marginal processes (see Cressie *et al.* (2010) Section 4) and second, considering temporal characteristics helps in source separation, that is determining which processes are causing the observed change. (ii) We extend the assessment to the whole of Antarctica and also consider an extra (fifth) process layer to explain fine-scale variation that is clearly affecting one of the data sets: satellite altimeter measurements of height change. (iii) To cater for the computational demands, which emerge at these scales and data volumes, we design an inference scheme that is parallelisable and demonstrate the potential speed-ups on a high-performance computer. As in Zammit-Mangion *et al.* (2014), the approach is sensitive to our prior beliefs, and we need to take certain justified modelling assumptions before running update schemes. We note that spatio-temporal inversion using a process-based framework such as that proposed here has been used elsewhere in studies of the cryosphere, for example by Hurkmans *et al.* (2012), and from a physical-statistical modelling perspective, by Berliner *et al.* (2008). However, these works concentrated on a small region such as a single outlet glacier or ice stream and do not consider an under-determined problem. For reference, the area of the Antarctic ice sheets exceeds 13 M km², and some processes operate over spatial scales of 20 km². This work is the first of its kind to provide an ice-sheet wide spatio-temporal inversion of all dominating ice sheet processes.

Using the work in Zammit-Mangion *et al.* (2014) as a foundation, the framework is constructed with standard hierarchical spatio-temporal modelling tools (Cressie and Wikle, 2011) in Section 2. In Section 3, we outline the use of numerical models to elicit informative priors and detail a parallelisable inference scheme to infer all other unknowns. In Section 4, we validate and visualise the results and compare them to other works. Section 5 concludes. Code used in this work is available in *R software* from the project website <http://www.rates-antarctica.net/>. A package and accompanying vignette facilitating most of the operations described in this work is available from the first author's *git* website <https://github.com/andrewzm/MVST>.

2. HIERARCHICAL MODELLING

We will follow a hierarchical modelling approach to the problem, as exemplified in works such as Berliner *et al.* (2000) and Wikle and Berliner (2007). The interested reader is referred to Cressie and Wikle (2011) for a comprehensive review. In summary, the model we will define is structured with (i) an observation layer detailing the interactions between the instruments and the processes under investigation; (ii) the process layer containing information on the spatio-temporal nature of the processes; and (iii) the parameter layer encoding prior beliefs on unknowns, which appear in any of the first two layers.

The nature of the observations in this application are sufficiently complex to warrant a split of the top layer into two sublayers. The first layer identifies interactions that may occur between the observations themselves. This is primarily included to cater for averaging effects occurring in the observations with large spatial footprint. The second layer describes the measurement mapping between the process layer and the instrument itself. The 'process' we consider is a superposition of a multivariate spatio-temporal model and a fine-scale effect. Although the fine-scale effect is not of direct interest, it is clearly present in the altimetric observations and thus needs to be considered.

In this work, we will adopt a hybrid approach where we take advantage of the geophysical numerical models to estimate some of the parameters and infer other parameters from data using a Markov chain Monte Carlo (MCMC) scheme. The fourth layer describes prior distributions over these latter parameters, which appear both in the first observation layer and in the component describing the fine-scale variation. A graphical model showing all parameters (both estimated and random), processes and data sets is given in Figure 1 for use as reference throughout this section.

2.1. Data preprocessing and the data model

There are, broadly, two satellite-based techniques used when assessing the SLR contribution of the Antarctic ice sheet. The first, satellite altimetry, is a technique for measuring ice-sheet height change, while the second, gravimetry, is a technique for measuring changes in the Earth's gravity field, the geoid, which can be used to infer mass changes. While the former produces point-referenced data, the latter is a potential field that generally produces large spatial footprints (often expressed as a spherical harmonic expansion). In this work, we made use of the most informative data sets available to date, from the

1. laser-based Ice Cloud and Elevation Satellite (ICESat, altimetry),
2. radar-based Environment Satellite (EnviSat, altimetry),
3. Gravity Recovery and Climate Experiment (GRACE, gravimetry).

In addition to these, we also use GPS data, which provide a direct reading of bedrock uplift rates.

The unit of interest is *height change* in m year⁻¹ due to each of the four processes of interest. We therefore preprocessed the data accordingly to obtain spatially distributed height or mass (in case of GRACE) rates and associated errors for each year. The altimetry data were processed on a 20×20 km grid as in Riva *et al.* (2009), while the GRACE observations were preprocessed as in Zammit-Mangion *et al.* (2014). The time interval for this study was restricted to the overlapping epochs of the operating instruments, namely 2003–2009. GRACE data for 2003 were, however, omitted on recommendation of Luthcke (personal communication) because of issues with data quality for this year in the product we were using. We also included pseudo point-referenced observations to weakly constrain the height rates due to ice, surface and firn processes to zero outside the continent's coast. These observations, of which we have just over 24 000, were placed on the triangulation vertices of the relevant processes (Section 2.4) in the ocean and were assigned a variance of 0.001. In all, we make use of just

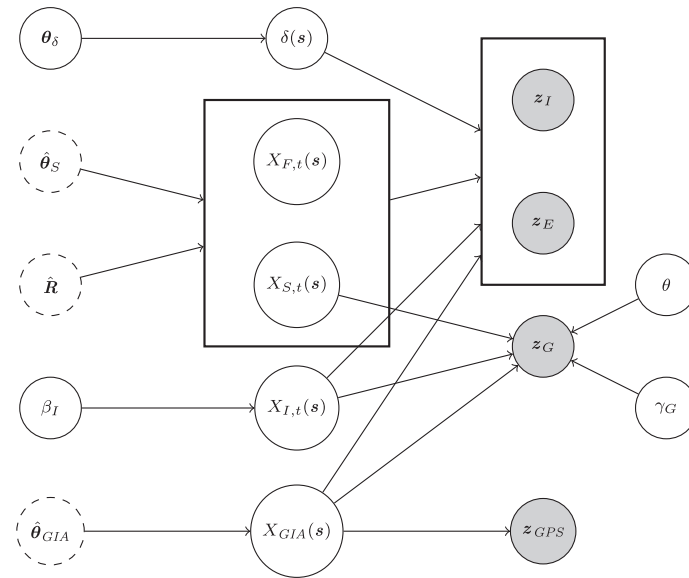


Figure 1. Graphical model for the problem under study. Variables on the left are parameters pertaining to the underlying spatio-temporal processes. θ_δ denotes the fine-scale variation ($\delta(s)$) parameters, $\hat{\theta}_S$, parameters relating to the surface and firm processes ($X_{S,t}(s)$ and $X_{F,t}(s)$), \hat{R} the interaction between these two processes, β_I the weightings of covariates associated with the model for height change due to ice dynamics ($X_{I,t}(s)$) and $\hat{\theta}_{GIA}$ parameters appearing in the model for glacio-isostatic adjustment ($X_{GIA}(s)$). Variables in the middle denote the spatio-temporal processes, while variables on the right denote the observations (filled circles). The parameters θ and γ_G are parameters appearing in the observation equation for z_G . Rectangles denote groupings: the altimetry data sets z_E and z_I are treated identically (same observation characteristics), while $X_{S,t}(s)$ and $X_{F,t}(s)$ are, jointly, a multivariate spatio-temporal process. Dashed circles enclose the parameters estimated using maximum likelihood from data/numerical models prior to computing the Bayesian update. The pseudo observations are omitted for clarity

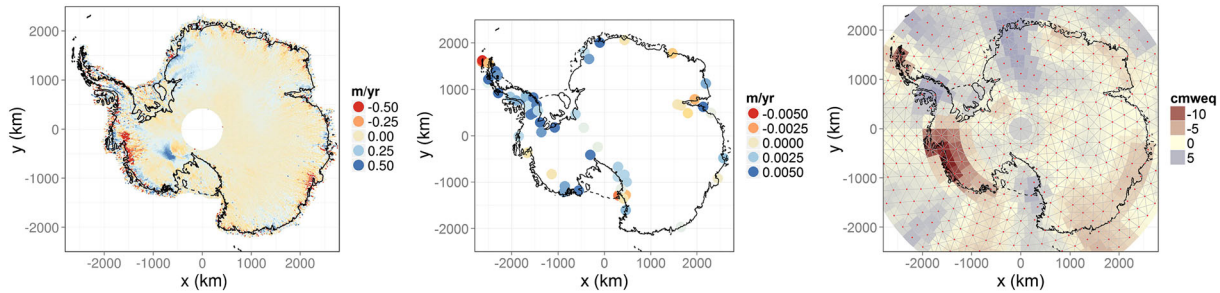


Figure 2. Snapshots from some of the data sets employed in this study. **Left:** Ice Cloud and Elevation Satellite elevation rates in 2006, **Centre:** global positioning system elevation rates in 2006 and **Right:** Gravity Recovery and Climate Experiment mass anomalies in 2006 in centimetres water equivalent (cmweq) with overlaid neighbourhood graph. The continent's grounding line and coastline are denoted using solid and dashed lines, respectively

over 365 000 observations, 2832 of which have a large spatial footprint (corresponding to the GRACE data). Snapshots from ICESat, GPS and GRACE in 2006 are shown in Figure 2, and the role of the data sets used (excluding the pseudo-data) is summarised in Table 1.

2.1.1. First data layer

We use the first data layer to describe the mapping between the observations and the processes. Denote the ICESat rates at year t as $z_{I,t}$, the EnviSat rates as $z_{E,t}$, the GRACE rates as $z_{G,t}$, the GPS rates as $z_{GPS,t}$ and let $z_t = [z_{I,t}^T, z_{E,t}^T, z_{G,t}^T, z_{GPS,t}^T]^T$. Further, denote the respective errors as $e_t = [e_{I,t}^T, e_{E,t}^T, e_{G,t}^T, e_{GPS,t}^T]^T$. Then, we can write the first data layer as

$$z_t = P(\theta)y_t + e_t, \quad t = 0 \dots T \quad (1)$$

where $y_t = [y_{I,t}^T, y_{E,t}^T, y_{G,t}^T, y_{GPS,t}^T]^T$ is defined in Section 2.1.2 and $P(\theta) = \text{bdiag}(I_I, I_E, P_G(\theta), I_{GPS})$, where $\text{bdiag}(\cdot)$ places its arguments along a block diagonal. We assume that the altimetry and GPS observations are point-referenced independent records of the first layer, hence the use of the identity matrix I_i . On the other hand, it is well known that the effective resolution of GRACE is less than that supplied (Luthcke *et al.*, 2013), and the role of $P_G(\theta)$ is to cater for this observation interaction. $P_G(\theta)$ is thus defined as

Table 1. Observation data sets employed, together with type of spatial footprint, the processes which they are able to detect changes in and the number of values available for the temporal horizon under consideration (2003–2009, GRACE 2004–2009)

Data	Footprint	Detects	Number	Symbol
ICESat	point	fine scale, surface, firm, ice processes and GIA	193 914	z_I
EnviSat	point	fine scale, surface, firm, ice processes and GIA	164 551	z_E
GRACE	polar polygons ($\approx 40\,000\text{ km}^2$)	surface, ice processes and GIA	2832	z_G
GPS	point	GIA	147	z_{GPS}

ICESat, Ice Cloud and Elevation Satellite; EnviSat, Environment Satellite; GRACE, Gravity Recovery and Climate Experiment; GPS, global positioning system; GIA, glacio-isostatic adjustment.

$$P_G^{(i,j)} = \begin{cases} \theta, & i \sim j, (\#i)\theta > 0.9 \\ 1 - (\#i)\theta & i = j, (\#i)\theta > 0.9 \\ \frac{1}{(\#i) + 1} & \text{otherwise} \end{cases} \quad (2)$$

where $(\#i)$ denotes the number of neighbours of observation i and \sim denotes ‘neighbour of’. Equation (2) describes the proportion of signal $(\#i)\theta$ that should be attributed to the spatial regions associated with the neighbouring observations. If $(\#i)\theta$ exceeds 0.9 (indicative of poor localisation), the observation is assumed to be an equal average of itself and its neighbours. Two GRACE observations are marked as neighbours if their geometric centres are distanced by less than 450 km, see Figure 2 (right).

For the GRACE signal, preprocessing techniques might also induce an under-estimation on the variance of $e_{G,t}$ (Zammit-Mangion *et al.*, 2014). To allow for this, we redefine $e_{G,t} := \gamma e_{G,t}^*$ where $e_{G,t}^*$ is the original supplied error and γ_G is an inflation parameter, which, together with θ (the strength of neighbourhood interaction), needs to be estimated.

2.1.2. Second data layer

The second data layer models the relationship between the processes, fine-scale variation $\delta_t(s)$ and the data. There are four dominating physical processes that need to be considered when estimating SLR contribution. (i) The first is a solid-Earth process known as glacio-isostatic adjustment (GIA), the response of the lithosphere to past changes in ice loading that took place, primarily, at the end of the last glacial, around 12 KYr ago; (ii) on top of the bedrock sits the ice, which constantly flows from the interior to the ocean; (iii) between the surface snow and the ice at depth lies a firm layer that has a density between that of snow and ice and that is constantly compacting; (iv) at the surface, the mass balance is affected mostly by changes in precipitation, sublimation and (limited) melt-water run-off. The role of these layers and the way they interact with the instrumentation are discussed in greater detail in Zammit-Mangion *et al.* (2014). In the current geophysical setting, the fine-scale variation is due to localised (fine-scale) precipitation anomalies (Isaksson and Karlen, 1994), which are most evident in areas of rapidly varying topography (Richardson *et al.*, 1997).

The model for the second layer is as follows

$$y_t = \mathcal{A}_{S,t} X_{S,t}(s) + \mathcal{A}_{F,t} X_{F,t}(s) + \mathcal{A}_{I,t} X_{I,t}(s) + \mathcal{A}_{GIA,t} X_{GIA}(s) + \mathcal{A}_{\delta,t} \delta_t(s) \quad (3)$$

where the $\{X_{i,t}\}$ denote the change in height (in m year^{-1}) due to processes that contain some spatio-temporal structure (as opposed to δ_t); the overall surface process anomaly (S), firm densification (F), ice dynamics/subshelf melting (I) and GIA (GIA), respectively. Note that because we are operating in a multivariate setting, we have not modelled fine-scale variation to be part of a particular process of interest (e.g. Katzfuss and Cressie, 2011 and Kang and Cressie, 2011), so that fine-scale effects from all processes, if present, are captured in $\delta_t(s)$.

Each element in the vectors of linear operators $\mathcal{A}_{i,t}$ maps the physical process $X_{i,t}(s)$ or the fine-scale variation into the observation space. Each element in $\mathcal{A}_{i,t}$ depends on the type and the location of the observation. For the data sets described earlier, it is sufficient to let each element in $\mathcal{A}_{i,t}$ (say $\mathcal{A}_{i,t}^j$) to be an integral operator of the form

$$\mathcal{A}_{i,t}^j X_{i,t}(s) = \int_{\Omega_j} f_{i,t}^j(s) X_{i,t}(s) ds \quad (4)$$

for $i \in \{S, F, I, GIA, \delta\}$, where Ω_j is the observation footprint, $f_{i,t}^j(s) \propto 1$ or $1/|\Omega_j|$ in the case of a purely accumulating or averaging observation, respectively, and $f_{i,t}^j(s) = \delta(s - s_{j,t})$ in the case of a point-referenced observation at location s_j at time t . For the altimetry data sets, pseudo-data sets and GPS, because their spatial support is relatively small, we assumed point-referenced observations and thus let $\Omega_j = \Omega$ (the entire domain) and $f_{i,t}^j(s) = \delta(s - s_{j,t})$ for all processes. For the GPS and pseudo observations, we let $f_{\delta,t}^j(s) = 0$ for each j because these observations are not affected by the fine-scale process.

For GRACE, we let $f_{i,t}^j(s) = \rho_i(s)$ for each $i \in \{I, S, F, GIA, \delta\}$, where each $\rho_i(s)$ is a spatially varying density map used to convert the height change to the detected mass change. $\rho_I(s) := \rho_I$ is the density of ice (917 kg m^{-3}), $\rho_S(s)$ is the density at which surface process

fluctuations occur (spatially varying between 300 and 600 kg m⁻³) obtained from the Regional Atmospheric Climate Model (RACMO, Lenaerts *et al.*, 2012), $\rho_F(s) := \rho_F = 0$ because firn compaction is a mass-preserving process and $\rho_{GIA}(s) := \rho_{GIA}$ is the effective mantle density (3800 kg m⁻³) (RACMO, used to set $\rho_S(s)$, is a regional climate model that reconstructs, at a resolution of 27 km, various surface processes from global climate reanalysis data. It will also be used in Section 3 to provide information on the spatial-temporal characteristics of the surface processes). For GRACE and $i \in \{S, F, I\}$, we also define Ω_j as the intersection between the observation footprint and the continent's grounding line, because height loss by all processes (except GIA) do not contribute to changes in the geoid beyond this line; see Zammit-Mangion *et al.* (2014) for details. We evaluate the integral for GRACE by fine gridding the footprint domain although Monte Carlo integration can also be used (Katzfuss and Cressie, 2011).

Rewriting $A_{\delta,t}^j \delta_t(s)$ as a Wiener integral, it is straightforward to show that the observed fine-scale variation at the j^{th} observation is distributed as

$$A_{\delta,t}^j \delta_t(s) \sim \mathcal{N}\left(\mathbf{0}, \int_{\Omega_j} |f_{\delta,t}^j(s)\sigma(s)|^2 ds\right) \quad (5)$$

where $\sigma(s)$ is the spatially varying variance of the uncorrelated process. For homogeneous fine-scale variation, when averaging over the footprint area, $\text{var}(A_{\delta,t}^j \delta_t(s))$ scales as $1/|\Omega_j|$ while when accumulating (as is the case with GRACE), as $|\Omega_j|$. The latter implies, at first sight, that this variation can have a large impact on the GRACE readings. However, placing realistic values for the fine-scale variation, which roughly correspond to the maximum *a posteriori* estimates in Section 3.2, yielded a variation that was three orders of magnitude lower than that of the GRACE signal itself ($\text{var}(e_G^*)$). Because of this parameter insensitivity, we omitted the effect of fine-scale variation on the GRACE readings by setting $f_{\delta,t}^j(s) = \mathbf{0}$ for these data. This assumption simplifies estimation of the fine-scale variation parameters that are, in any case, mostly informed through the altimetry data.

2.2. The process layer

The third layer describes the evolution of the multivariate spatio-temporal model. Three of the variates are spatio-temporal, while the fourth, GIA, is spatial (repeatedly observed at each time point). For all the three spatio-temporal processes, we assume a spatio-temporal auto-regressive process of the general form

$$X_{i,t}(s) = a_i(s)X_{i,t-1}(s) + d_i^T \beta(s) + w_i(s) \quad (6)$$

where $a_i(s)$ is a possibly spatially varying auto-regressive parameter, $w_i(s) \sim \mathcal{GP}(0, k_i(s, r; \theta_i))$, where $\mathcal{GP}(\mu(\cdot), k(\cdot, \cdot))$ denotes a Gaussian process (Rasmussen and Williams, 2006) with mean $\mu(\cdot)$ and kernel $k(\cdot, \cdot)$, where k is not necessarily spatially invariant, $d_i^T = [d_{1,t}, d_{2,t}, \dots]$ are covariates and $\beta(s)$ the respective spatial weightings, which need to be estimated. We now discuss the models for each of the spatio-temporal processes in turn.

2.2.1. Ice dynamics/subshelf melting

Height changes due to ice dynamics, and subshelf melting is assumed to be relatively smooth over the short time scales considered, that is they follow a linear trend. This assumption follows from studies that show approximately linear temporal variation in ice discharge (Rignot *et al.*, 2011b). The ensuing model for this process is

$$X_{I,t}(s) = [1 \quad t] \beta_I(s) + w_{I,t}(s) \quad (7)$$

where $\beta_I(s) \in \mathbb{R}^2$ represents the spatially varying mean change and accelerating change in height, respectively, due to ice dynamics. We assume that both $\beta_I(s)$ and the disturbance $w_{I,t}(s)$ have small spatial scales and are assumed spatially uncorrelated following discretisation. Prior distributions for $\beta_I(s)$ are given in Section 2.3. Fully specifying the statistics of $w_{I,t}$ is crucial to avoid confounding with the fine-scale variation (which is mostly due to precipitation anomalies). We thus take $\text{cov}(w_{I,t}(s), w_{I,t'}(s)) = 0.01^2$ (i.e. we allow for confounding on scales of 1 cm year⁻¹ if $t = t'$ and 0 otherwise. This choice also reflects the prior assumed (temporal) smoothness of height change due to ice dynamics and subshelf melting.

2.2.2. Surface processes

Changes in height due to surface processes fluctuate rapidly at large (ice sheet) scales (Rignot *et al.*, 2011b). To check whether the same holds at small (20 km grid) scales, we fitted an auto-regressive(1) model to each spatial point from anomalies with respect to the 30-year mean for the output of the forward model RACMO. The auto-regressive coefficient was found to be positive in some places and negative in others. This variation, in addition to spatially varying amplitudes and length scales evident from RACMO, suggested that a heterogeneous model needs to be fitted to this process. As in Zammit-Mangion *et al.* (2014), we thus adopted the technique of fitting local homogeneous models on a coarsely gridded domain and then allowing the parameters to vary smoothly between the grid cells.

Locally, for each coarse-gridded cell, we assume that the surface process follows the local interactions model

$$X_{S,t}(s) = a_S X_{S,t-1}(s) + w_{S,t}(s) \quad (8)$$

where a_S serves to define a random-walk space-time process (Stroud *et al.*, 2001; Zammit-Mangion *et al.*, 2012). Note that this is the auto-regressive interpretation of a space-time Gaussian field with separable covariance function $k(u, \Delta) = k_1(u)k_2(\Delta)$ where $u = \|s - r\|$, $\Delta = |t - t'|$ and $k_2(\Delta) = a_S^\Delta$. It can be shown (Storvik *et al.*, 2002) that in this case $\text{var}(w_{S,t}(s)) = (1 - a_S^2)k_1(u)$.

For each cell, we let $k_1(\cdot)$ (the spatial covariance function) be a Matérn function given by

$$k_1(u) = \frac{\sigma_S^2}{2^{\nu_S-1}\Gamma(\nu_S)}(\kappa_S u)K_{\nu_S}(\kappa_S u) \quad (9)$$

where σ_S^2 is the marginal variance, κ_S is the scaling parameter, ν_S is the smoothness parameter and K_{ν_S} is the modified Bessel function of the second kind. Use of the Matérn is motivated by the dimensionality reduction methods employed in Section 2.4. The unknown parameters across all gridded cells are collected in $\theta_S = (a_S^T, \kappa_S^T, \sigma_S^2, \nu_S)^T$.

2.2.3. Firm processes

Changes in height due to surface and firm processes are highly anticorrelated (more precipitation implies more compaction and *vice versa*). We capture this interaction by extending the spatial coregionalisation approach in Finley *et al.* (2007); Zammit-Mangion *et al.* (2014) to its spatio-temporal equivalent, termed the spatio-temporal linear model of coregionalisation by Choi *et al.* (2009). As in the spatial case, we model $X_{F,t}(s)$ as

$$X_{F,t}(s) = bX_{S,t} + c\tilde{X}_{1,t} \quad (10)$$

where $\tilde{X}_{1,t}(s)$ is independent but identically distributed as $X_{S,t}(s)$, that is

$$\tilde{X}_{1,t} \sim \mathcal{GP}(0, k_1(\cdot)k_2(\cdot)) \quad (11)$$

We can thus express the joint distribution as

$$(X_{S,t}(s), X_{F,t}(s)) \sim \mathcal{GP}(0, \mathbf{R} \otimes [k_1(\cdot)k_2(\cdot)]), \quad \mathbf{R} = \begin{bmatrix} 1 & b \\ b & b^2 + c^2 \end{bmatrix} \quad (12)$$

where \mathbf{R} is a 2×2 correlation matrix composed of scalars b and c defining the point-wise interactions. Note that this definition of \mathbf{R} is a scaled version of that in Zammit-Mangion *et al.* (2014) because here the marginal variance of the process is included in the definition of $k_1(\cdot)$.

It can be shown, by comparing second moments, that the Gaussian field (12) is expressed in auto-regressive form as

$$\begin{bmatrix} X_{S,t}(s) \\ X_{F,t}(s) \end{bmatrix} = \begin{bmatrix} a_S \\ a_S \end{bmatrix} \begin{bmatrix} X_{S,t-1}(s) \\ X_{F,t-1}(s) \end{bmatrix} + \begin{bmatrix} w_S(s) \\ w_F(s) \end{bmatrix} \quad (13)$$

where

$$(w_S(s), w_F(s)) \sim \mathcal{GP}(0, (1 - a_S^2)[\mathbf{R} \otimes k_1(\cdot)]) \quad (14)$$

It is apparent here that coregionalisation does not induce off-diagonal elements in the propagation matrix. Interactions are induced solely through the correlated added disturbances. \mathbf{R} is unknown and needs to be estimated.

2.2.4. Glacio-isostatic adjustment

Glacio-isostatic adjustment is a slowly varying process with time scales on the order of thousands of years. For the 7-year period considered, we thus model it as a spatial process (which is repeatedly observed at each time point):

$$X_{GIA} \sim \mathcal{GP}(0, k_{GIA}(u; \theta_{GIA})) \quad (15)$$

where k_{GIA} is a Matérn kernel, parameterised by some unknown parameter vector $\theta_{GIA} = (\kappa_{GIA}, \sigma_{GIA}^2, \nu_{GIA})^T$.

2.2.5. Fine-scale variation

We model the fine-scale variation $\delta_t(s)$ as a stochastic volatility model as in Katzfuss and Cressie (2012)

$$\delta_t(s) \sim \mathcal{N}(\mathbf{0}, \exp(\eta_1 + b(s)\eta_2)) \quad (16)$$

where $b(s)$ is a basis function used to explain some of the fine-scale variation. Because we envision fine-scale variation to be present in high-relief regions, we use terrain roughness as a basis function $b(s)$. The roughness function was derived from a high-pass channel image from the Moderate Resolution Imaging Spectroradiometer Mosaic of Antarctica Image map at 750 m resolution (Haran *et al.*, 2005). The image was binned in 20 km boxes and the spread (standard deviation) of the image intensity in each box found. The logarithm of this spread incremented by 1 (to have a minimum roughness of 0), seen in Figure 3, left, was then used as the basis function for fine-scale variation. For identifiability reasons, we assumed that the statistics of fine-scale variation are temporally invariant. The parameters $\theta_\delta = [\eta_1, \eta_2]^T$ are unknown and need to be estimated.

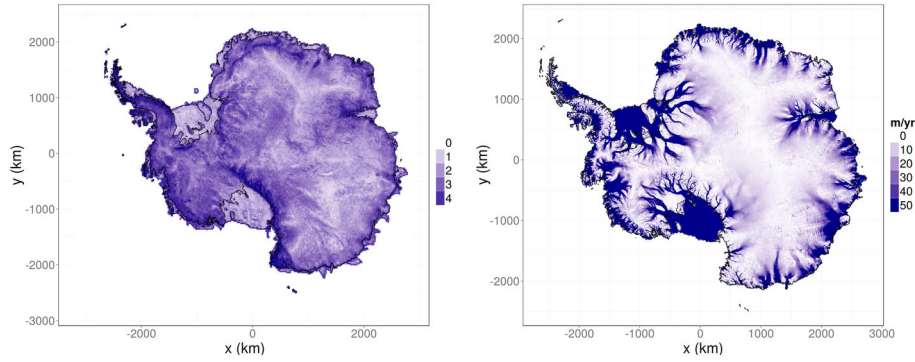


Figure 3. **Left:** Basis function used for small-scale variation, derived as described in main text. **Right:** Surface velocity map, used to shrink height change due to ice dynamics towards zero in areas of low velocity

2.3. The parameter layer

Because of identifiability concerns, we adopt a hybrid approach to parameter estimation, in which some process parameters, $\Theta_1 = (\mathbf{R}, \theta_{GLA}, \theta_S)$, are estimated from numerical models as in Calder *et al.* (2011), and the rest, $\Theta_2 = (\theta_\delta, \gamma^2, \theta, \beta_I(s))$, are estimated from the data in an MCMC scheme. The estimation procedure for the former group is described in Section 3.1. For the latter, we apply the following prior beliefs:

- $\theta_\delta = [\eta_1, \eta_2]^T$: These parameters describe the second-order statistics of the fine-scale variation and can take values on the real line. For this reason, we let

$$\eta_1 \sim \mathcal{N}(\mu_{\eta_1}, \sigma_{\eta_1}^2) \quad (17a)$$

$$\eta_2 \sim \mathcal{N}(\mu_{\eta_2}, \sigma_{\eta_2}^2) \quad (17b)$$

where the hyper-parameters $(\mu_{\eta_1}, \mu_{\eta_2}, \sigma_{\eta_1}^2, \sigma_{\eta_2}^2)$ are set following a preliminary study employing a simplified process model and only one of the data sets (Section 3.2).

- θ : This parameter describes the extent of averaging within a neighbourhood of GRACE observations and should be positive. However, from (2), observe that \mathbf{P}_G is invariant to θ when $(\#i\theta)$ exceeds 0.9, which occurs at $\theta \approx 0.14$ for the observation arrangement shown in Figure 2. For this reason, we constrain $\theta < 0.14$ by equipping it with a uniform prior

$$\theta \sim \mathcal{U}(0, \theta_p) \quad (18)$$

where \mathcal{U} denotes the uniform distribution and $\theta_p = 0.14$.

- γ_G : This parameter allows for variance inflation of the GRACE signal. To obtain a tractable conditional distribution, we let

$$\gamma_G^2 \sim IG(\alpha_{\gamma_G}, \beta_{\gamma_G}) \quad (19)$$

where IG is the inverse-Gamma distribution and the hyper-parameters are configured such that the 95 percentile $q_{0.95} = 400$ and the 5 percentile $q_{0.05} = 1$, reflecting our belief that we do not expect the variance inflation to be outside the range $[1, 400]$. Suitable values are $\alpha_{\gamma_G} = 0.59$ and $\beta_{\gamma_G} = 2.15$.

- $\beta_I(s)$: The spatially distributed parameters $\beta_1(s), \beta_2(s)$ define the linear and accelerating change in height due to ice sheet processes, respectively. The marginal variances of these are set to be high for regions where the ice is flowing fast, and low (with a shrinkage to zero) otherwise, to reflect the fact that height change due to ice dynamics at very low speeds ($< 10 \text{ m s}^{-1}$) is unlikely. We thus let

$$\sigma_i(s) = \frac{\sigma_{sat,i}}{1 + \exp(-(v(s) - 10))}, \quad i = 1, 2 \quad (20)$$

where we set $\sigma_{sat,1} = 15 \text{ m year}^{-1}$, $\sigma_{sat,2} = 0.5 \text{ m year}^{-2}$ for one sector of Antarctica, the Amundsen Sea Embayment (which is known to exhibit large accelerations, see Scott *et al.*, 2009) and $\sigma_{sat,2} = 0.1 \text{ m year}^{-2}$ elsewhere. Here, $v(s)$ is the ice velocity obtained from synthetic aperture radar data (Rignot *et al.*, 2011a), supplemented with balance velocities (Bamber *et al.*, 2000) where data are missing. The reconstructed velocity map used is shown in Figure 3, right. Note that although we have a spatially continuous specification for these parameters, they will be estimated following discretisation (together with the processes) on a triangulation within a Bayesian setting, see Section 2.4.

2.4. Gaussian field approximation

While convenient for modelling, Gaussian field spatio-temporal process models are not suitable for large-scale inference because the computational complexity of a required Bayesian update scales as $O(m^3)$ where m is the number of data points. The dependence on m^3

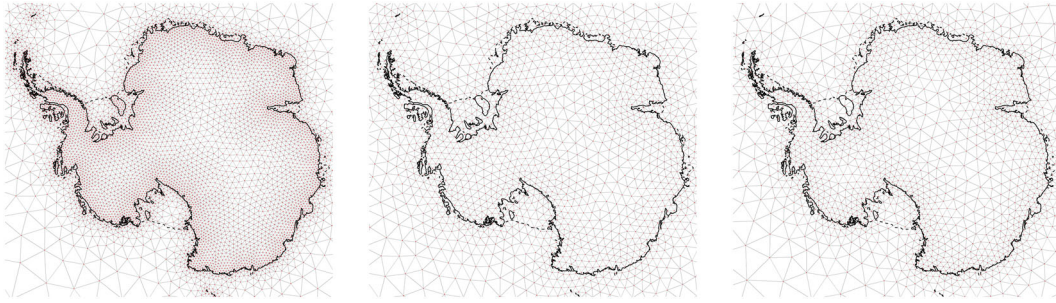


Figure 4. Meshes used to reconstruct the spatio-temporal processes. From left to right: (i) $X_{I,t}(s), \beta_{I,1}(s), \beta_{I,2}(s)$; (ii) $X_{S,t}(s), X_{F,t}(s)$; and (iii) $X_{GIA}(s)$

is prohibitive in several applications, including the one considered here. One remedy is to project functions from the original, infinite-dimensional space \mathcal{O} onto some lower-dimensional space \mathcal{O}_n through a set of basis functions $\{\phi_i\}_{i \in \mathbb{N}}$. A natural way to approximate Gaussian fields is by noting that if these are of the Matérn type, they can be treated as solutions to stochastic partial differential equations (SPDEs) of the Laplace type (Lindgren *et al.*, 2011). It can be shown that the projection onto \mathcal{O}_n of such an SPDE with compactly supported basis functions yields a Gaussian Markov Random Field (GMRF) with favourable computational properties (Rue and Held, 2005). The basis functions we employ are ‘tent’ functions constructed on the triangulations shown in Figure 4.

The GMRFs are obtained by decomposing the spatio-temporal field at each time point and the spatially distributed parameters $(\beta_1(s), \beta_2(s))$ as a weighted sum of the tent basis functions, that is we let $X_{i,t}(s) \approx \phi_i(s)^T x_{i,t}$ and $\beta_{I,i}(s) \approx \phi_i(s)^T \beta_{I,i}$. This decomposition allows us to rewrite the second data layer as

$$y_t = C_t x_t + \delta_t \quad (21)$$

where $C_t = [\mathcal{A}_{S,t} \phi_S(s)^T, \mathcal{A}_{F,t} \phi_F(s)^T, \mathcal{A}_{I,t} \phi_I(s)^T, \mathcal{A}_{GIA,t} \phi_{GIA}(s)^T]$, $\delta_t = \mathcal{A}_{\delta,t} \delta_t(s)$ and the multivariate spatio-temporal model (ignoring GIA for now) is

$$\begin{bmatrix} x_{S,t} \\ x_{F,t} \\ x_{I,t} \end{bmatrix} = \begin{bmatrix} \mathcal{A}_S \\ \mathcal{A}_F \\ \mathbf{I} \end{bmatrix} \begin{bmatrix} x_{S,t-1} \\ x_{F,t-1} \\ x_{I,t-1} \end{bmatrix} + \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{I} \end{bmatrix} \begin{bmatrix} \beta_{I,1} \\ \beta_{I,2} \end{bmatrix} + \begin{bmatrix} w_{S,t} \\ w_{F,t} \\ w_{I,t} \end{bmatrix} \quad (22)$$

where \mathbf{I} is the identity matrix of appropriate size. Here, $\mathcal{A}_S \equiv \mathcal{A}_F$ describe the temporal propagation of the processes and are diagonal under the local separability condition assumed. $\beta_{I,1}$ and $\beta_{I,2}$ are discretised versions of $\beta_{I,1}(s)$ and $\beta_{I,2}(s)$, and $w_{i,t} \sim \mathcal{N}_c(\mathbf{0}, \mathbf{Q}_{w,i})$ is added random forcing, where $\mathcal{N}_c(\mathbf{h}, \mathbf{Q})$ denotes the normal distribution with canonical parameters \mathbf{h} and \mathbf{Q} . For the surface field, $\mathbf{Q}_{w,S}$ (which describes heterogeneous fields) is obtained by assuming that $w_{S,t}$ is the discretised solution to an SPDE with spatially varying parameters, as discussed in detail in Zammit-Mangion *et al.* (2014) and Lindgren *et al.* (2011) Section 3.2. The approach has the neat property that, *locally*, $w_{S,t}$ can be interpreted as a discretised solution to a standard Matérn field. Details on how we obtain the spatially distributed parameters from RACMO are given in Appendix A.1. Following coregionalisation between the surface and firn fields, jointly we have that

$$\begin{bmatrix} w_{S,t} \\ w_{F,t} \end{bmatrix} \sim \mathcal{N}_c(\mathbf{0}, [\mathbf{R}^{-1} \otimes \mathbf{Q}_{w,S}]) \quad (23)$$

Finally, as discussed in 2.2.1, we set $\mathbf{Q}_{w,I}$ diagonal, with precisions given by $1/0.01^2$.

We next combine the auto-regressive processes by defining $\mathbf{x}_t = [x_{S,t}^T, x_{F,t}^T, x_{I,t}^T]^T$ and $\mathbf{w}_t \sim \mathcal{N}_c(\mathbf{0}, \mathbf{Q}_w)$. The multivariate process can then be written in block form as

$$\mathbf{x}_t = \mathbf{A} \mathbf{x}_{t-1} + \mathbf{D}_t \boldsymbol{\beta}_I + \mathbf{w}_t \quad (24)$$

The full process describes Gaussian trajectories, so that we can represent the third layer as a GMRF over $(\mathbf{x}, \boldsymbol{\beta}_I)$ where $\mathbf{x} = [x_0^T, \dots, x_T^T, x_{GIA}^T]^T$. Let $\boldsymbol{\Theta} = \{\boldsymbol{\Theta}_1, \boldsymbol{\Theta}_2^{\beta_I}\}$, that is the set of all unknown parameters except $\boldsymbol{\beta}_I$. Then,

$$(\mathbf{x}, \boldsymbol{\beta}_I | \boldsymbol{\Theta}) \sim \mathcal{N}_c(\mathbf{0}, \mathbf{Q}) \quad (25)$$

The full precision structure \mathbf{Q} is obtained by considering the joint latent process, which decomposes as follows:

$$p(\mathbf{x}, \boldsymbol{\beta}_I | \boldsymbol{\Theta}) = p(\mathbf{x}_{0:T}, \mathbf{x}_{GIA}, \boldsymbol{\beta}_I | \boldsymbol{\Theta}) \quad (26a)$$

$$= p(\mathbf{x}_{0:T} | \mathbf{x}_{GIA}, \boldsymbol{\beta}_I, \boldsymbol{\Theta}) p(\mathbf{x}_{GIA} | \boldsymbol{\beta}_I, \boldsymbol{\Theta}) p(\boldsymbol{\beta}_I | \boldsymbol{\Theta}) \quad (26b)$$

$$= p(\mathbf{x}_{0:T} | \boldsymbol{\beta}_I, \boldsymbol{\Theta}) p(\mathbf{x}_{GIA} | \boldsymbol{\Theta}) p(\boldsymbol{\beta}_I | \boldsymbol{\Theta}) \quad (26c)$$

$$= \left(\prod_{i=1}^T p(\mathbf{x}_i | \mathbf{x}_{i-1}, \boldsymbol{\beta}_I, \boldsymbol{\Theta}) \right) p(\mathbf{x}_0 | \boldsymbol{\beta}_I, \boldsymbol{\Theta}) p(\boldsymbol{\beta}_I) p(\mathbf{x}_{GIA} | \boldsymbol{\Theta}) \quad (26d)$$

where (26c) follows from (i) $\mathbf{x}_{0:T}$ being conditionally independent of \mathbf{x}_{GLA} given $(\boldsymbol{\beta}_I, \boldsymbol{\Theta})$ and (ii) \mathbf{x}_{GLA} being conditionally independent of $\boldsymbol{\beta}_I$ given $\boldsymbol{\Theta}$. Equation 26d follows from $\boldsymbol{\beta}_I$ being independent of $\boldsymbol{\Theta}$. The prior $p(\boldsymbol{\beta}_I) = \mathcal{N}_c(\mathbf{0}, \mathbf{Q}_\beta)$ (where \mathbf{Q}_β follows directly from (20)), while the initial process distribution is defined as

$$\mathbf{x}_0 | \boldsymbol{\beta}_I, \boldsymbol{\Theta} \sim \mathcal{N}(\mathbf{D}_0 \boldsymbol{\beta}_I, [\mathbf{Q}_w - \mathbf{A}^T \mathbf{Q}_w \mathbf{A}]^{-1}) \quad (27)$$

We choose the precision $\mathbf{Q}_0 := \mathbf{Q}_w - \mathbf{A}^T \mathbf{Q}_w \mathbf{A}$, and not that of the respective stationary distribution, as it is sparse for sparse \mathbf{Q}_w and diagonal \mathbf{A} , thus preserving computational tractability. Further, the implied covariance matrix is representative for diagonal \mathbf{A} and \mathbf{Q}_w ; in particular, if $\mathbf{A} = c\mathbf{I}$ for $c \in (-1, 1)$, the induced covariance is the same as that of the stationary distribution (see Hamilton, 1994, Section 10.2).

It can be shown that for this model the full precision matrix over $(\mathbf{x}, \boldsymbol{\beta}_I)$ is

$$\mathbf{Q} = \begin{bmatrix} \mathbf{Q}_1 & \mathbf{Q}_2 \\ \mathbf{Q}_2^T & \mathbf{Q}_3 \end{bmatrix} \quad (28)$$

where \mathbf{Q}_1 is the precision matrix without exogenous inputs (see Rue and Held, 2005, Ch. 1 for a univariate equivalent)

$$\mathbf{Q}_1 = \begin{bmatrix} \mathbf{A}^T \mathbf{Q}_w \mathbf{A} + \mathbf{Q}_0 & -\mathbf{A}^T \mathbf{Q}_w & & & \\ -\mathbf{Q}_w \mathbf{A} & \mathbf{A}^T \mathbf{Q}_w \mathbf{A} + \mathbf{Q}_w & -\mathbf{A}^T \mathbf{Q}_w & & \\ & -\mathbf{Q}_w \mathbf{A} & \mathbf{A}^T \mathbf{Q}_w \mathbf{A} + \mathbf{Q}_w & -\mathbf{A}^T \mathbf{Q}_w & \\ & & \ddots & \ddots & \\ & & & \mathbf{A}^T \mathbf{Q}_w \mathbf{A} + \mathbf{Q}_w & -\mathbf{A}^T \mathbf{Q}_w \\ & & & & \mathbf{Q}_w \end{bmatrix} \quad (29)$$

and $\mathbf{Q}_2, \mathbf{Q}_3$ are defined as

$$\mathbf{Q}_2 = \begin{bmatrix} \mathbf{A}^T \mathbf{Q}_w \mathbf{D}_1 - \mathbf{Q}_0 \mathbf{D}_0 \\ \mathbf{A}^T \mathbf{Q}_w \mathbf{D}_2 - \mathbf{Q}_w \mathbf{D}_1 \\ \mathbf{A}^T \mathbf{Q}_w \mathbf{D}_3 - \mathbf{Q}_w \mathbf{D}_2 \\ \vdots \\ -\mathbf{Q}_w \mathbf{D}_T \end{bmatrix}, \quad \mathbf{Q}_3 = \mathbf{D}_0^T \mathbf{Q}_0 \mathbf{D}_0 + \sum_{i=1}^T (\mathbf{D}_i^T \mathbf{Q}_w \mathbf{D}_i) + \mathbf{Q}_\beta \quad (30)$$

Because of the tractability of the linear state-space model, we have combined $\boldsymbol{\beta}_I$ with \mathbf{x} , although the former are ‘parameters’ in the conventional sense. Because we have represented $(\mathbf{x}, \boldsymbol{\beta}_I)$ jointly, by concatenating across time we can rewrite the observation and process layers as follows:

$$\mathbf{z} = \mathbf{P}(\boldsymbol{\theta})\mathbf{y} + \mathbf{e} \quad (31a)$$

$$\mathbf{y} = \mathbf{C}'\mathbf{x}' + \boldsymbol{\delta} \quad (31b)$$

$$\mathbf{x}' | \boldsymbol{\Theta} \sim \mathcal{N}_c(\mathbf{0}, \mathbf{Q}) \quad (31c)$$

where $\mathbf{C}' = [\mathbf{C}, \mathbf{0}]$, $\mathbf{x}' = [\mathbf{x}^T, \boldsymbol{\beta}_I^T]^T$.

3. PARAMETER ESTIMATION AND INFERENCE

3.1. Estimation of parameters from geophysical models

Because we have three classes of data (altimetry, gravimetry and GPS) and four processes to infer, the problem is under-determined, rendering process-related parameter estimation from observations particularly difficult. We hence make use of numerical (geophysical) models for estimating the process parameters, namely $\boldsymbol{\Theta}_1$. Specifically, denote the vector of physical observations as \mathbf{z} and numerical model output as \mathbf{x}_M (we use the symbol \mathbf{x} to emphasise that we treat numerical model outputs as *directly observed* data). \mathbf{x}_M is the output of a deterministic model $g(\cdot)$ run with suitably defined parameters $\boldsymbol{\psi}$ and boundary conditions BC . The first assumption we make is that the true (unobserved) process \mathbf{x} and $\mathbf{x}_M = g(\boldsymbol{\psi}, BC)$ are conditionally independent, which we model as $p(\mathbf{x}, \mathbf{x}_M | \boldsymbol{\theta}) = p(\mathbf{x} | \boldsymbol{\theta})p(\mathbf{x}_M | \boldsymbol{\theta})$ for some parameter $\boldsymbol{\theta} \in \boldsymbol{\Theta}_1$, see Figure 5.

The second assumption is that \mathbf{x}_M is highly informative on $\boldsymbol{\theta}$ so that we can safely restrict the posterior parameter density to a point estimate: $p(\boldsymbol{\theta} | \mathbf{x}_M) = \delta(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}(\mathbf{x}_M))$. Then,

$$\begin{aligned} p(\mathbf{x} | \mathbf{x}_M, \mathbf{z}) &\propto p(\mathbf{z} | \mathbf{x}) \int p(\mathbf{x} | \boldsymbol{\theta}) p(\boldsymbol{\theta} | \mathbf{x}_M) d\boldsymbol{\theta} \\ &= p(\mathbf{z} | \mathbf{x}) p(\mathbf{x} | \hat{\boldsymbol{\theta}}(\mathbf{x}_M)) \end{aligned} \quad (32)$$

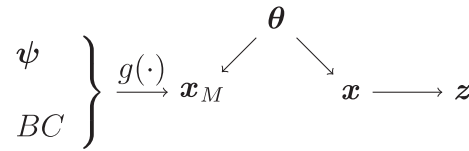


Figure 5. Assumed relationship between the geophysical (deterministic) model output $x_M = g(\psi, BC)$, the true process x , the observations z and underlying parameters θ . Conditioned on θ , x_M and x are assumed to be independent. The observations z are recordings of the true process x

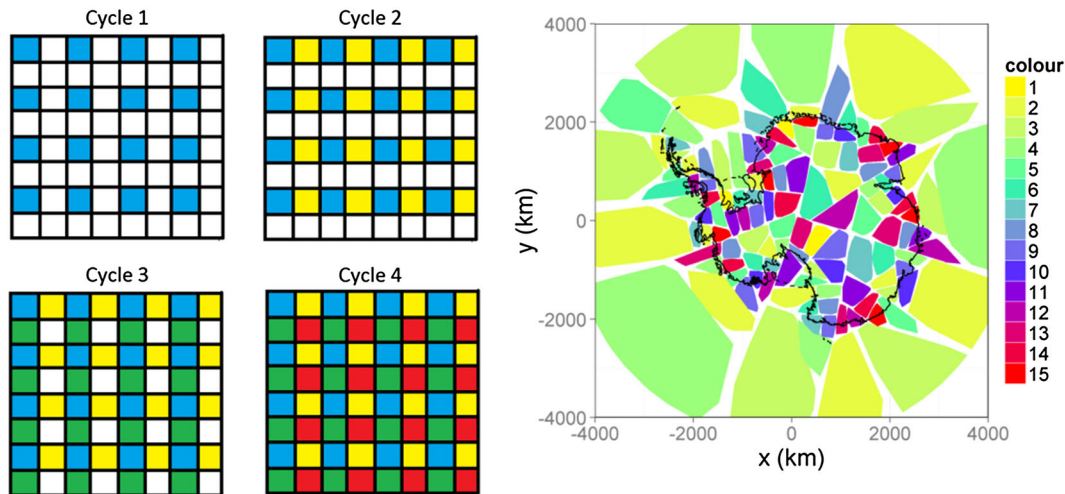


Figure 6. Left: Dispatching blocks under an 8×8 raster blocking scheme. A complete sample is obtained after the fourth cycle. Right: The final partitioning after applying the Kernighan–Lin algorithm seven times and a greedy colouring algorithm

Equation (32) suggests a two-stage approach to inference. First, find the maximum likelihood estimate of θ under x_M : $\hat{\theta}(x_M)$. Second, carry out a Bayesian update on x using z and $\hat{\theta}(x_M)$ as plug-in values in $p(x|\theta)$. This approach is similar to that shown in Calder *et al.* (2011). Details for this first stage as applied to the present application are given, for the relevant processes, in Appendix A.

3.2. Markov chain Monte Carlo implementation

In order to carry out inference over x and Θ_2 , we need to complete the model specification by eliciting prior beliefs over θ_δ . We do this by using information from the log-likelihood of these parameters, evaluated at several points on a much simplified model. The log-likelihood reveals that η_1 and η_2 are highly correlated because of the large flat surface in the continent's interior; we therefore also transform these parameters to aid mixing. Details for this step are given in Appendix B.

The sheer size of the problem under study devotes special attention. The combined dimensionality of (x, β_I) is over 80 000, we have over 300 000 observations, crucially some of which have a large spatial support. The time taken to compute the Cholesky factor of Q following an approximate minimum degree permutation, and the marginal variance using the *SuiteSparse* package (Davis, 2011), required about 5 h on a mid-end server. This is prohibitive even for the vanilla implementation of the popular integrated nested Laplace approximations (INLA, Rue and Martino, 2009) package. We deal with this by implementing a scheme that can be parallelised and implemented on a distributed system.

We divide the state-space x' into d contiguous spatial blocks as in Zammit-Mangion *et al.* (2014). Denote the indices of the i th subset of x', x'^i , as \mathcal{I}_i and let $\mathcal{I} = \bigcup_i \mathcal{I}_i$. The conditional posterior distribution for the i th block is given by $x'^i \sim \mathcal{N}_c(\hat{Q}^{ii} \hat{x}'^i, \hat{Q}^{ii})$ where

$$\hat{Q}^{ii} = C'^{iT} Q_{e\delta} C'^i + Q^{ii} \quad (33a)$$

$$\hat{Q}^{ii} \hat{x}'^i = C'^{iT} Q_{e\delta} \left(z - \sum_{j \in [1 \dots d]/i} C'^j \hat{x}'^j \right) - \sum_{j \in [1 \dots d]/i} (Q^{ij} \hat{x}'^j) \quad (33b)$$

where Q^{ij} , $i, j \in [1 \dots d]$ is the matrix composed of the rows in Q with indices in \mathcal{I}_i and the columns of Q with indices in \mathcal{I}_j and where C'^i is the matrix composed of the columns in C' with column indices in \mathcal{I}_i . Here, $Q_{e\delta} = [Q_e^{-1} + Q_\delta^{-1}]^{-1}$, $Q_e = \text{prec}(e)$ and $Q_\delta = \text{prec}(\delta)$.

In this work, we extend the approach by blocking with a view to parallelisation over a multinode architecture. Parallelisation is possible in Gibbs sampling if one can find groups of $\{x'^i\}$ that are independent when conditioned on the rest of the state space and the data. Groups because groups within this set, say \mathcal{J}^j , can then be updated in any order desired (or, indeed, even in parallel) for a given deterministic sweep strategy. As an example, consider the following scenario, depicted in Figure 6, left. Here, we have divided the entire state space into a raster of 8×8 blocks. Assume each block is conditionally independent of every other block given the neighbouring blocks and the data, then we

can obtain a sample of \mathbf{x}' by sampling all nonadjacent blocks in parallel. For this simple case, we only need four cycles, one for each block set $\mathcal{J}^1 \dots \mathcal{J}^4$; a potential speed-up of $64/4 = 16$ is achievable. More generally, a simple parallel Gibbs sampling proceeds as follows: first, instruct computational nodes to sample from blocks with index in \mathcal{J}^1 in parallel, wait for the cycle to complete, update and distribute the latest sample and repeat for $\mathcal{J}^2, \dots, \mathcal{J}^{n_j}$. This sampling strategy is known as chromatic sampling (Gonzalez *et al.*, 2011), because the assigning of conditional independent groups is akin to the problem of graph colouring (Molloy and Reed, 2002): the problem of finding a colour arrangement such that no two neighbours in a graph have the same colour.

We cannot employ a simple raster blocking method and a four-colouring scheme (which is possible for all graphs on a plane, see Gonthier, 2008) for two reasons. First, the key drawback with raster partitioning is the unequal load distribution when employing variable mesh density (as in this problem). Blocks of equal size will, in general, contain drastically different numbers of vertices. This, in turn, will cause computer nodes in a high-performance environment to 'wait' unnecessarily for the block containing the largest number of vertices to be completed before the next cycle can begin. We remedy this problem by dividing the graph into subgraphs of roughly equal size using the *Kernighan–Lin* algorithm (Kernighan and Lin, 1970). In this study, we use the implementation by Ladroue (2011) seven times in order to subdivide the graph into 128 roughly equal-sized partitions. The final partitioning scheme we used in our problem is shown in Figure 6 (right): with this blocking strategy, we obtained a median block size of 609 vertices with a minimum of 590 and a maximum of 750, which is indicative of a good load balance.

The second problem is that, *conditioned on the data*, the adjacency graph over the blocks is no longer a sufficient constraint for conditional independence. This is made apparent from the second term within the brackets of (33b). This term implies that because of dense rows in \mathbf{C}' , $\hat{\mathbf{x}}^i$ may even depend on states that are in blocks not immediately adjacent to block i . Thus, when colouring the graph, in addition to the constraint that no two adjacent blocks have the same colour, we have to apply the constraint that no observation can be informative of more than one block of a given colour. This is particularly relevant to the present problem because GRACE has a support that spans multiple blocks. We coloured the graph using a greedy algorithm on a breadth-first-search ordering of the vertices. The final arrangement with the lowest amount of colours, 15 in this case, is shown in Figure 6, right.

We grouped the GIA field on its own (i.e. assigned it a colour of its own) because it is highly spatially correlated. With 1832 vertices, the GIA block is about three times larger than any other block so that the theoretical maximum speed-up is thus $129/(15 + 3) = 7.2$. In practise, we saw a fivefold speed-up over use of a single thread when sampling the state space only, which is remarkably close given the communication overhead, notable in this case because of message passing of the updated samples. Sampling of all the other parameters ($\Theta_2^{\beta_I}$, see Section 3.3) needs to be done sequentially; the final speed-up we obtained for one whole MCMC cycle was of about three; however, this is still significant given the time needed to generate one sample. Moreover, this approach extends nicely to larger state spaces, if required. We generated 16 000 samples on a high-performance computer (having 16 2.6 GHz SandyBridge cores per node and 4 GB memory per core) using four nodes and 24 parallel threads with Rmpi in 42 h.

3.3. Sampling the parameters and diagnostics

Following an updated sample of \mathbf{x} and β_I , we sample from the conditionals of the remaining parameters, defined as follows

$$p(\theta | \mathbf{z}_G, \mathbf{x}', \gamma) \propto \begin{cases} \exp(-\frac{1}{2\gamma^2} (\mathbf{z}_G - \mathbf{P}_G(\theta) \mathbf{C}'_G \mathbf{x}')^T \mathbf{Q}_{eG} (\mathbf{z}_G - \mathbf{P}_G(\theta) \mathbf{C}'_G \mathbf{x}')) & \theta \in [0, \theta_p] \\ 0 & \text{otherwise} \end{cases} \quad (34a)$$

$$p(\gamma | \mathbf{z}_G, \mathbf{x}', \theta) \propto \exp \left(-\frac{1}{2\gamma^2} (\mathbf{z}_G - \mathbf{P}_G(\theta) \mathbf{C}'_G \mathbf{x}')^T \mathbf{Q}_{eG} (\mathbf{z}_G - \mathbf{P}_G(\theta) \mathbf{C}'_G \mathbf{x}') - \frac{n_G}{2} \ln \gamma^2 - (\alpha_\gamma + 1) \ln \gamma^2 - \beta_\gamma / \gamma^2 \right) \quad (34b)$$

$$p(\theta_\delta | \mathbf{z}, \mathbf{x}') \propto \exp \left(-\frac{1}{2} (\mathbf{z}_A - \mathbf{C}'_A \mathbf{x}')^T \mathbf{Q}_{e\delta} (\mathbf{z}_A - \mathbf{C}'_A \mathbf{x}') - \frac{1}{2\sigma_{\eta_1}^2} (\eta_1 - \mu_{\eta_1})^2 - \frac{1}{2\sigma_{\eta_2}^2} (\eta_2 - \mu_{\eta_2})^2 \right) \quad (34c)$$

where n_G is the number of GRACE observations, $\mathbf{Q}_{eG} = \text{prec}(\mathbf{e}^*)$ (recall that \mathbf{e}^* is the error prior to inflation), \mathbf{C}'_G is the submatrix of \mathbf{C}' corresponding to the GRACE observations and \mathbf{C}'_A is that corresponding to the altimetry (ICESat and EnviSat). Note that (34c) is independent of θ and γ as a result of the assumption $f_s^j(s) = \mathbf{0}$ for the GRACE observations in (5).

From this note that only γ^2 can be sampled from directly. For θ and θ_δ , we initially carried out a Metropolis-within-Gibbs step by proposing from a Gaussian distribution centred at zero with variance 0.1^2 . However, the resulting effective sample size from this scheme was remarkably low, prompting us to use univariate slice sampling instead (Neal, 2003), which considerably decreased autocorrelation. We implemented the stepping-out methods and adapted the interval width used for the first 50 samples, following which the interval was fixed. These samples were then discarded.

Visual inspection of the traceplots confirm good mixing of the chains. Geweke diagnostics and autocorrelation plots on these traces also support convergence. To assess convergence in \mathbf{x} and β_I , we took 2000 variables at random from \mathbf{x}' and carried out Geweke diagnostics on each of the traces. The z scores from the diagnostics were then verified to be normally distributed (expected under the assumption of trace independence, a reasonable one given the continent's size and correlation lengths). Geweke diagnostics were carried out using the R package coda.

4. VALIDATION, VISUALISATION AND RESULTS

In this section, we summarise the main inferential results. In Section 4.1, we test the model using an independent data set and use the results to expose some limitations of the approach, while in Section 4.2, we focus on the marginal fields. Finally, in Section 4.3, we concentrate on

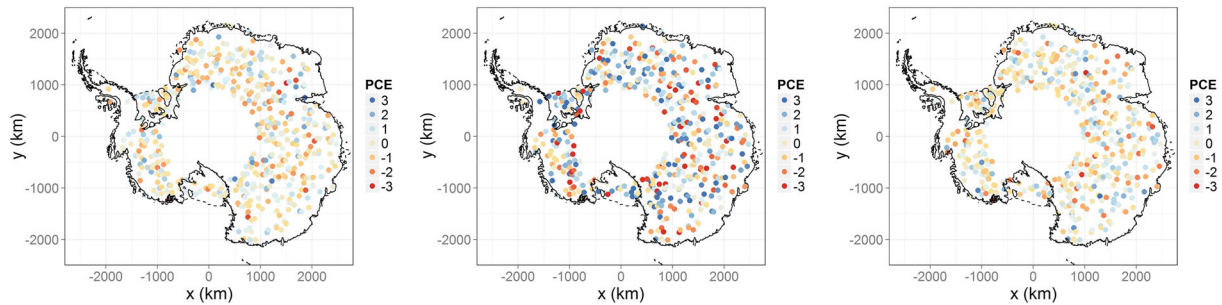


Figure 7. Pivoted Cholesky residuals (PCE) of validation data using real data (centre) and pseudo-data generated from the posterior distribution (left and right)

mass balance estimates, that is we determine where and by how much Antarctica is losing or gaining mass. In all Antarctica plots presenting our results, we use bilinear interpolation to reconstruct the mesh surface, overlay the triangulation to convey the native process resolution and stipple with green dots where the absolute posterior mean is greater than one posterior standard deviation.

4.1. Diagnostic check using a separate remote sensing data set with pivoted Cholesky residuals

In this section, we check the model using data from an altimeter on EnviSat's predecessor, the European Remote-sensing satellite-2 (ERS-2). Because of process smoothness, we compute residuals that are 'corrected' for correlation through the pivoted Cholesky residuals, which are derived from the covariance matrix of the residuals (Bastos and O'Hagan, 2009). The 'pivot' in the pivoted Cholesky decomposition of the matrix sorts the residuals as follows: the first element is that with the highest variance, the second element is the largest variance conditioned on the first element and so on. In a spatio-temporal setting, locations in the vicinity of each other should be far apart in the ordering, even if they have similar marginal variances, that is the variance of the first point in a close pair should be low when conditioned on the second point if correlations are correctly captured. The advantage of this approach is that proximity and spatio-temporal correlation are taken into account when visualising the residuals. As such little correlation should be visually apparent when inspecting plots of the pivoted Cholesky residuals (as opposed to when visualising the standardised residuals). We carry out this diagnostic check by choosing 500 validation points from the ERS-2 data. The pivoted Cholesky residuals are seen in the centre panel of Figure 7 where only some correlation is apparent, indicating that the chosen spatial structure of the model is not implausible.

A pivoted Cholesky residual plot can be put into perspective by carrying out a 'Turing test'. This involves sampling validation data points (at the space-time locations) from the predictive distribution, re-evaluating the pivoted Cholesky residuals for this pseudo data and comparing the resulting plots to that of the true data. This informal test checks whether we can spot the real-data case from the pseudo-data case. Overall, from Figure 7, we find it relatively easy to pick out the pseudo-data cases from the real-world case, the latter exhibiting more extreme residuals, representative of overconfidence. A density plot of the pivoted Cholesky residuals (not shown) indicates that the residuals are centred at zero but also confirms an overall overconfidence. It is not clear whether this is a result of model mis-specification or an issue with the residuals in the ERS-2 data set; however, it does hint at the possibility of our posterior estimates being slightly overconfident.

4.2. Posterior fields

In Figure 8, we show results for the mass rates due to ice dynamics for the 7 years under consideration together with uncertainty estimates. These results corroborate the spatial results of Zammit-Mangion *et al.* (2014) where only West Antarctica was considered. The most prominent features include the negative, accelerating rates in the Pine Island Glacier and Thwaites Glacier (inset) and an overall ice build-up (with no significant acceleration) in the Kamb area (inset). In East Antarctica, several expected features are also apparent, most notably the considerable mass loss occurring in the Totten Glacier (Velicogna and Wahr, 2013, inset) and areas surrounding the Frost glacier (Rignot, 2006), both of which are grounded below sea level and are hence considered potentially unstable (for basin definitions and landmarks, see Figure 9). Some glaciers, such as Shirase glacier towards the top of the plot (reported to be largely in equilibrium by Pattyn and Derauw, 2002)), are shown to be thickening. We note that glacier thickening at large rates is physically improbable (except in the Kamb area, Ng and Conway, 2004) and that here there could be confounding between changes due to ice dynamics and persistent positive local precipitation anomalies. We omit changing height thickness over the ice shelves in the plots, as this is not a contributor to SLR.

In Figure 10, we provide a comparison of posterior estimates for height changes due to surface processes and those provided by the regional climate model, RACMO. When considering aggregates (right), the positive-negative trend is clearly followed by RACMO although, for East Antarctica, a bias is evident, possibly because of a systematic underestimation of the snowfall anomaly by RACMO for 2004–2007 (van Wessem *et al.*, 2014). It should be noted that East Antarctica is one of the most poorly sampled places on the planet in terms of meteorological observations and therefore also relatively poorly constrained in RACMO. Although in most cases outside our 95% credibility intervals, we find it encouraging that the shape and phase of the anomaly patterns could be reconstructed. Essentially, this result shows that RACMO reconstructs the temporally high frequency, spatially mid-frequency signals evident in satellite altimetry data but with differences in the amplitude of the signal.

In Figure 11, we provide our reproduction for the GIA field and compare it to a forward numerical model, IJ05R2, which we used to set the parameters on x_{GIA} . Notably, as in IJ05R2, we observe a positive trend (uplift) of 2–4 mm year⁻¹ GIA over West Antarctica and an overall negative signal (subsidence) towards the pole and over much of East Antarctica. The main discrepancy apparent in Figure 11 (right)

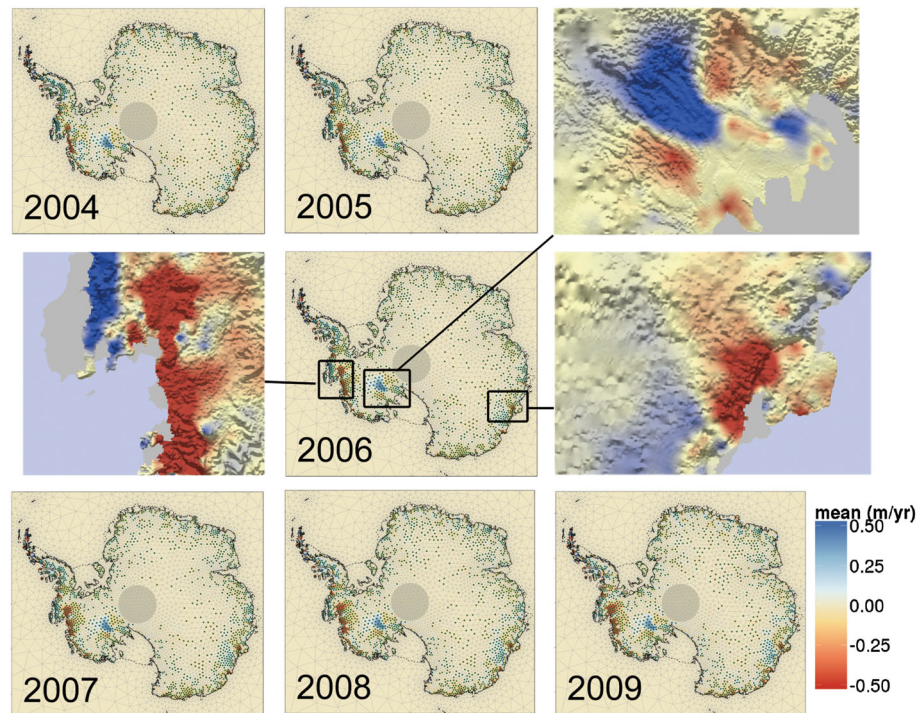


Figure 8. Evolution of height rates due to changes in ice dynamics from 2004 to 2009. Insets show close ups of the Pine Island Glacier (centre left), the Kamb Ice Stream (top right) and the Totten Glacier (centre right) in 2006, overlaid over the bedrock topography extracted from Bedmap2 (Fretwell *et al.*, 2013). Height changes over ice shelves (enclosed between the dashed and solid lines and grayed out in the insets) are omitted as these do not contribute to sea-level change

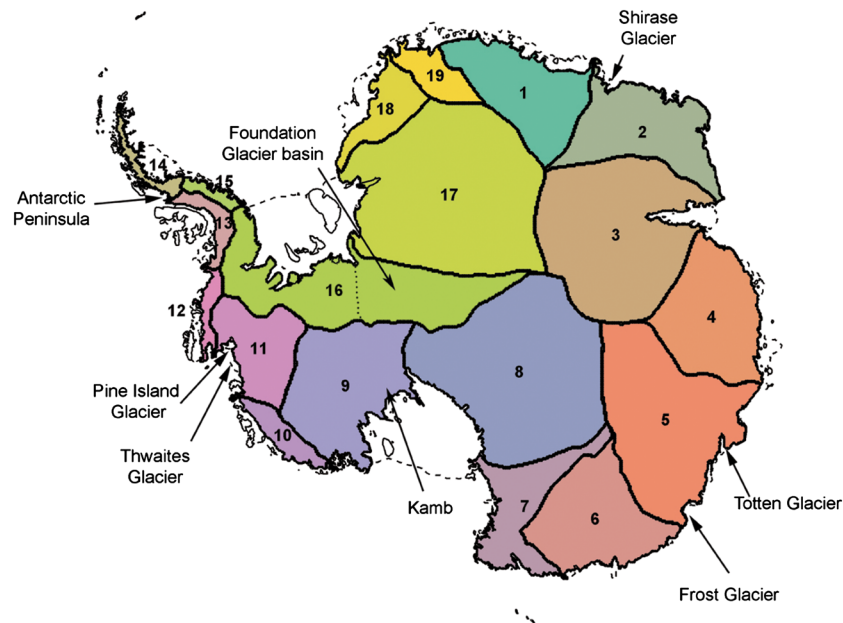


Figure 9. Basin locations corresponding to Rignot *et al.* (2011a) and landmarks. The West Antarctic ice sheet is composed of Basins 9, 10, 11, 12 and part of 16, the East Antarctic ice sheet of Basins 1–8, 17–19 and part of 16 and the Antarctic Peninsula of Basins 13–15. The Foundation Glacier basin (east part of Basin 16) is associated with the East Antarctic ice sheet (Section 4.3)

is the large positive rate observed in our results in basins 1, 17–19, which correspond to the upper part (in the figure) of the East Antarctic ice sheet. We caution that GPS readings used here have not been corrected for the elastic response of the lithosphere and that we are using release 1 of the GRACE mass-concentration data set (work on a second release is underway). Consequently, the GIA results (and, to a lesser extent, the SLR contribution estimates of Section 4.3) are not definitive and presented as illustrative of the framework capability.

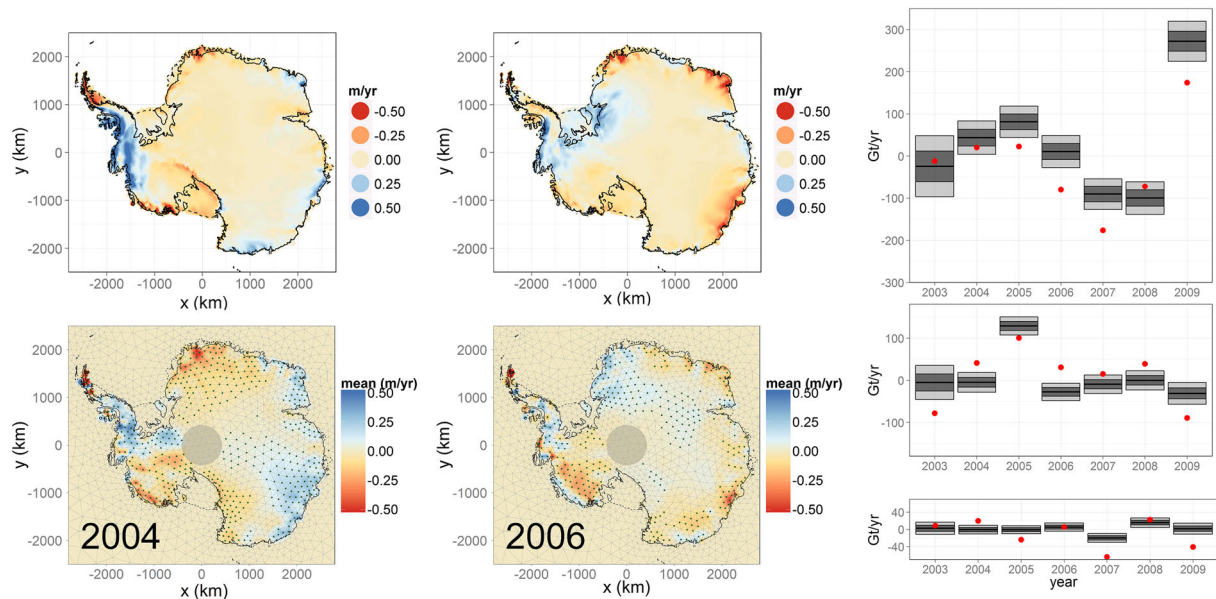


Figure 10. Left: Anomalies from a 30-year mean in the output of regional climate model (top) compared to the estimated surface process height rates (bottom) from the framework for 2004 and 2006. Right: Estimated rate of mass change due to surface processes (black bars) and the values given by the regional climate model (red dots) for the East Antarctic ice sheet (top), West Antarctic ice sheet (centre) and Antarctic Peninsula (bottom). The 1σ and 2σ levels are denoted by the dark and light shadings, respectively

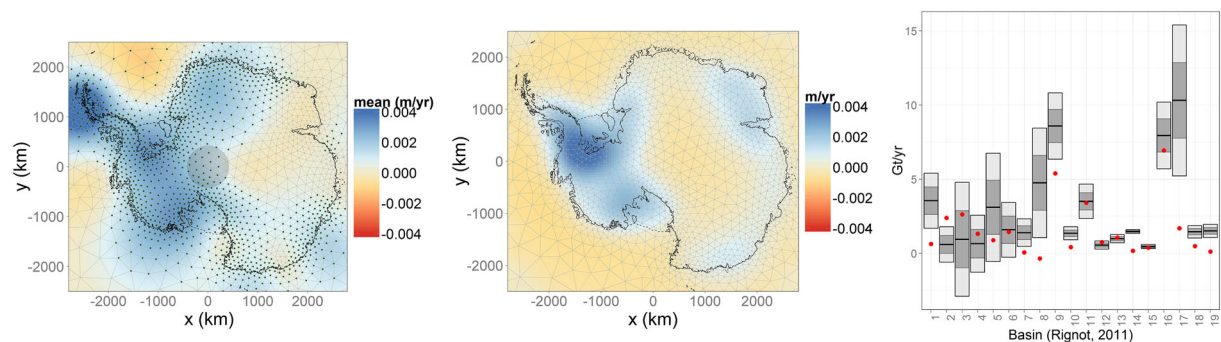


Figure 11. Left: Posterior estimate of x_{GIA} . Stipples mark the states where $|\mu| > 1\sigma$. Centre: The glacio-isostatic adjustment (GIA) model output of Ivins *et al.* (2013), IJ05R2. Right: The estimated total change of mass due to GIA on a basin scale (black bars) and the value given by IJ05R2 (red dots). The 1σ and 2σ levels are denoted by the dark and light shadings, respectively

4.3. Estimates of sea-level rise contribution

Multivariate spatio-temporal modelling allows us to discriminate between the processes, in order to obtain mass-balance estimates from assumed density profiles, the same profiles that are used to translate the observed height changes to the mass changes observed by GRACE. As discussed in Zammit-Mangion *et al.* (2014), several authors have provided SLR contribution estimates for Antarctica. Here, we compare to the GRACE-based method of Luthcke *et al.* (2013) (where the IJ05R2 model is subtracted from the GRACE signal) and the so-called input–output (IO) method, where mass balance is found by deducting the rate of ice discharge (obtained through satellite readings of horizontal ice velocity) from the surface mass balance, obtained from a regional climate model, in this case RACMO. We obtained these latter results from Shepherd *et al.* (2012), which are similar to those of Rignot *et al.* (2011a) but updated to include more recent data, particularly for measured ice thickness. We note that the mass imbalance is not strictly equivalent to SLR contribution when accounting for grounding line migration (Rignot *et al.*, 2011a), and thus, we can expect results from the IO method to be slightly more negative (although it is unclear whether the IO results had been corrected for this or not). To use similar basin definitions as these studies, we have allocated the Foundation Glacier basin to the East Antarctic ice sheet, see Figure 9.

Estimates of SLR contribution for the West, East Antarctic ice sheets and the Antarctic Peninsula are given in Figure 12. Again, there is a broad agreement between the results. In the West Antarctic ice sheet, the positive anomaly in 2005 is clear for the three methods, as is the apparent negative trend in the latter years of the study. For the East Antarctic ice sheet, the conclusions are similar, in that the overall positive-negative pattern along the years is preserved, despite our solutions being temporally more smooth, probably because of the imposed auto-regression coefficients on the surface process prior which are mostly positive. The pattern we observe in the Antarctic Peninsula is reproduced in the other solutions; however, the large mass loss indicated in the other methods is not within our credibility intervals. This

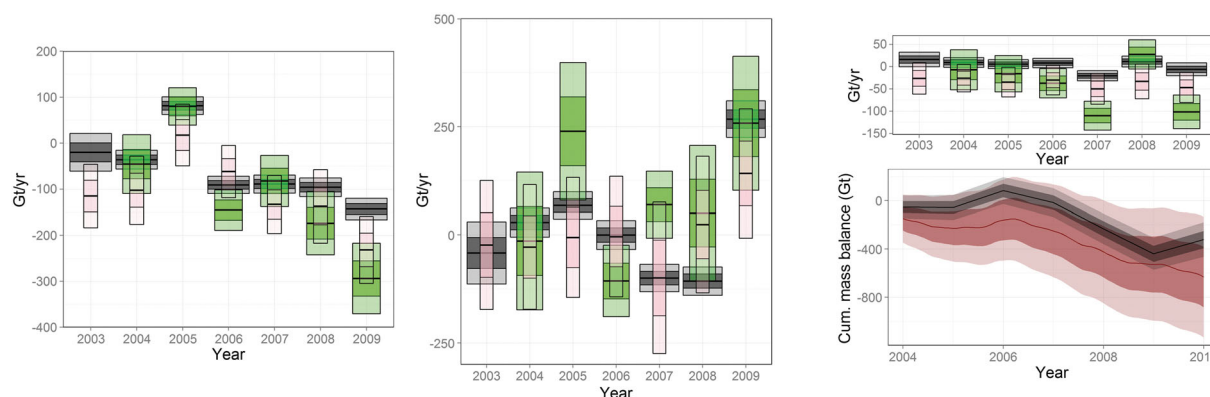


Figure 12. Mass-balance estimates (black) compared to those from the Gravity Recovery and Climate Experiment-based method of Luthcke *et al.* (2013) (green) and the IO method (pink) as reported in Shepherd *et al.* (2012). The 1σ and 2σ levels are denoted by the dark and light shadings, respectively. **Left:** West Antarctic ice sheet. **Centre:** East Antarctic ice sheet. **Top right:** Antarctic Peninsula. **Bottom right:** Estimated cumulative mass change throughout the period of study (black) compared to the estimate obtained when averaging over multiple methods (Shepherd *et al.*, 2012).

could be due to an optimistic prior belief in the definition over β_I or a result of inaccurate representation of the field by the chosen mesh. It could also be due to overestimation of mass loss in the other methods for a region that is particularly challenging for both GRACE and the IO method because of the small scale at which ice dynamic takes place.

Finally, we compare the cumulative mass change obtained using our method with Shepherd *et al.* (2012), which is an unweighted mean estimate (using a meta-analysis) from three approaches (GRACE-based methods, IO-methods and altimetry + RACMO methods) and multiple data preprocessing methods. Our cumulative mass change follows the unweighted mean estimate but is less negative. We note, however, that a more recent version of the RACMO model, v2.3, produces a more positive snowfall rate by over 100 Gt year^{-1} compared with the version used in Shepherd *et al.* (2012) (van Wessem *et al.*, 2014). This difference would increase the integrated IO method results by the same amount, resulting in a lower mass loss rate for the unweighted mean in Shepherd *et al.* (2012). Our estimated SLR contribution of Antarctica during this period (2003–2009) is of $0.13 \pm 0.08 \text{ mm year}^{-1}$. This lies, as expected, between the ICESat-based estimate of $-0.06 \pm 0.23 \text{ mm year}^{-1}$ and the GRACE-based estimate of $0.16 \pm 0.14 \text{ mm year}^{-1}$ as reported in Shepherd *et al.* (2012) between October 2003 and December 2008; however, it is considerably less than the GRACE-based method of Luthcke *et al.* (2013), which reported $0.23 \pm 0.07 \text{ mm year}^{-1}$ for 2003–2010. The discrepancy with the latter is possibly due to the inclusion of 2010, a year in which the West Antarctic ice sheet has continued to witness increased ice loss (McMillan *et al.*, 2014).

5. CONCLUSION

In this paper, we have shown that spatio-temporal modelling is a powerful tool in assessing Antarctica's mass balance and contribution to SLR. Unlike other methods, its use of forward numerical models is up to the assumption of conditional independence (used only to estimate parameters) and is thus less restrictive and bias-prone than standard data assimilation procedures. Overall, we find that our results conform quite well with those making explicit use of numerical models, indicating that this approach may be used in the future to validate the models with remote sensing data. This is particularly useful in Antarctica where *in situ* data are difficult and sometimes impossible to obtain.

There are several ways in which the framework can be improved. (i) More work needs to be done on mesh generation, with special care taken in high-relief and narrow areas such as the North Antarctic Peninsula. A coarse mesh, in these regions, constitutes an inaccurate model representation, which may be the cause of the optimistic results obtained in this region. (ii) A comparison with the auxiliary data set from ERS-2 revealed that the fields are unbiased and that the modelled spatial correlation is suitable. However, an overall overconfidence was observed, which might be the cause of the relatively narrow credibility intervals seen in Section 4.2. This could be a result of model mis-specification, but it is also possible that the supplied errors with the preprocessed data are, in fact, optimistic. (iii) Process parameters are treated as fixed and assumed to be fully informed by geophysical models. In an MCMC scheme, parameter uncertainty can be easily included, although this would require the provision of posterior densities over the unknown parameters. To remedy this, one could still assume conditional independence but not restrict the posterior parameter density to a point estimate as we do here. (iv) The assumption of imposed (local) spatio-temporal stationarity might be a fair assumption for some of the processes considered here; however, this does not hold for ice dynamics, which is directional and exhibits spatio-temporal interactions (Hurkmans *et al.*, 2012). It would be possible within the current framework to consider a simplified transport model for this process (e.g. Wikle *et al.*, 2001) although this will come at some computational expense.

The key novelties of this work lie in the level of complexity and sheer size of the spatio-temporal model employed, the way in which numerical models were incorporated with large-scale multivariate spatio-temporal systems and the tools developed for inference over multiple-process architectures. Because the work is multifaceted, we have implemented an R package MVST that facilitates most operations described in this paper and provides a simplified analysis of that described here as a vignette. One notable useful feature of this package is the ease with which multiple observations, with spatially varying observation operator, can be considered with various types of multivariate spatio-temporal systems, including those decomposed on triangulations such as those considered here. It is envisioned that this package can

be used to help set-up latent Gaussian multivariate spatio-temporal models in a variety of settings. Once the Gaussian model is set-up, it can be passed on for inference with other packages, such as INLA, although we also provide code for the elementary Gaussian case.

Acknowledgements

This work was made possible by NERC grant NE/I027401/1. The authors would like to thank Botond Cseke, Doug Nychka, Chris Wikle, Christophe Andrieu, Ruud Hurkmans, Bert Wouters, Noel Cressie, Volker Klemann and Katherine Rose for helpful discussions and tips throughout the project; Callum Wright and Gethin Williams for help with BlueCrystal; and Scott Luthcke, Frédérique Remy, Thomas Flament, Matt King and Liz Petrie for provision of and support with the data.

REFERENCES

- Bamber JL, Vaughan DG, Joughin I. 2000. Widespread complex flow in the interior of the Antarctic ice sheet. *Science* **287**(5456):1248–1250.
- Bastos LS, O'Hagan A. 2009. Diagnostics for Gaussian process emulators. *Technometrics* **51**(4):425–438.
- Berliner LM, Cressie N, Jezek K, Kim Y, Lam CQ, van der Veen CJ. 2008. Equilibrium dynamics of ice streams: a Bayesian statistical analysis. *Statistical Methods and Applications* **17**(2):145–165.
- Berliner LM, Wikle CK, Cressie N. 2000. Long-lead prediction of Pacific SSTs via Bayesian dynamic modeling. *Journal of Climate* **13**(22):3953–3968.
- Calder CA, Berrett C, Shi T, Xiao N, Munroe DK. 2011. Modeling space–time dynamics of aerosols using satellite data and atmospheric transport model output. *Journal of Agricultural, Biological, and Environmental Statistics* **16**(4):495–512.
- Choi J, Reich BJ, Fuentes M, Davis JM. 2009. Multivariate spatial-temporal modeling and prediction of speciated fine particles. *Journal of Statistical Theory and Practice* **3**(2):407–418.
- Cressie N, Shi T, Kang EL. 2010. Fixed rank filtering for spatio-temporal data. *Journal of Computational and Graphical Statistics* **19**(3):724–745.
- Cressie N, Wikle CK. 2011. *Statistics for Spatio-Temporal Data*. Wiley: Hoboken, NJ.
- Davis T: SPARSEINV: a MATLAB toolbox for computing the sparse inverse subset using the Takahashi equations. <http://www.cise.ufl.edu/research/sparse/SuiteSparse>, Online: Last accessed 28 April 2014.
- Finley AO, Banerjee S, Carlin BP. 2007. spBayes: an R package for univariate and multivariate hierarchical point-referenced spatial models. *Journal of Statistical Software* **19**(4):1–24.
- Fretwell P, Pritchard HD, Vaughan DG, Bamber JL, Barrand NE, Bell R, Bianchi C, Bingham RG, Blankenship DD, Casassa G, et al. 2013. Bedmap2: improved ice bed, surface and thickness datasets for Antarctica. *The Cryosphere* **7**:375–393.
- Gonthier G. 2008. Formal proof—the four-color theorem. *Notices of the AMS* **55**(11):1382–1393.
- Gonzalez J, Low Y, Gretton A, Guestrin C. 2011. Parallel Gibbs sampling: from colored fields to thin junction trees. *International Conference on Artificial Intelligence and Statistics*, Fort Lauderdale, FL, USA, 324–332.
- Gunter BC, Didova O, Riva REM, Ligtenberg SRM, Lenaerts J T M, King M A, van den Broeke M R, Urban T. 2014. Empirical estimation of present-day Antarctic glacial isostatic adjustment and ice mass change. *The Cryosphere* **8**:743–760.
- Hamilton JD. 1994. *Time Series Analysis*. Princeton University Press: Princeton.
- Haran TJ, Bohlander TS, Fahnestock M: MODIS mosaic of Antarctica (MOA) image map. National Snow and Ice Data Center, Boulder, Colorado USA: National Snow and Ice Data Center. <http://dx.doi.org/10.7265/N5ZK5DM5>, Online: Last accessed 28 April 2014.
- Hurkmans RTWL, Bamber JL, Sørensen LS, Joughin IR, Davis CH, Krabill WB. 2012. Spatiotemporal interpolation of elevation changes derived from satellite altimetry for Jakobshavn Isbræ, Greenland. *Journal of Geophysical Research* **117**(F3):F03001.
- Isaksson E, Karlen W. 1994. Spatial and temporal patterns in snow accumulation, Western Dronning Maud Land, Antarctica. *Journal of Glaciology* **40**(135):399–409.
- Ivins ER, James TS. 2005. Antarctic glacial isostatic adjustment: a new assessment. *Antarctic Science* **17**(04):541–553.
- Ivins ER, James TS, Wahr J, Schrama O, Ernst J, Landerer FW, Simon KM. 2013. Antarctic contribution to sea level rise observed by GRACE with improved GIA correction. *Journal of Geophysical Research: Solid Earth* **118**(6):3126–3141.
- Kang EL, Cressie N. 2011. Bayesian inference for the spatial random effects model. *Journal of the American Statistical Association* **106**(495):972–983.
- Katzfuss M, Cressie N. 2011. Spatio-temporal smoothing and EM estimation for massive remote-sensing data sets. *Journal of Time Series Analysis* **32**(4):430–446.
- Katzfuss M, Cressie N. 2012. Bayesian hierarchical spatio-temporal smoothing for very large datasets. *Environmetrics* **23**(1):94–107.
- Kernighan BW, Lin S. 1970. An efficient heuristic procedure for partitioning graphs. *The Bell System Technical Journal* **49**(2):291–307.
- Ladroue C: Graph bisection in R. <http://www.r-bloggers.com/graph-bisection-in-r/>, Online: Last accessed 28 April 2014.
- Lenaerts JTM, Van den Broeke MR, van de Berg WJ, van Meijgaard E, Munneke PK. 2012. A new, high-resolution surface mass balance map of Antarctica (1979–2010) based on regional atmospheric climate modeling. *Geophysical Research Letters* **39**(4):L04501.
- Ligtenberg SRM, Helsen MM, van den Broeke MR. 2011. An improved semi-empirical model for the densification of Antarctic firn. *The Cryosphere* **5**(4):809–819.
- Lindgren F, Rue H, Lindström J. 2011. An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. *Journal of the Royal Statistical Society B* **73**(4):423–498.
- Luthcke SB, Sabaka TJ, Loomis BD, Arendt A, McCarthy JJ, Camp J. 2013. Antarctica, Greenland and Gulf of Alaska land-ice evolution from an iterated GRACE global mascon solution. *Journal of Glaciology* **59**(216):613–631.
- McMillan M, Shepherd A, Sundal A, Briggs K, Muir A, Ridout A, Hogg A, Wingham D. 2014. Increased ice losses from Antarctica detected by CryoSat-2. *Geophysical Research Letters* **41**(11):3899–3905.
- Molloy M, Reed B. 2002. *Graph Colouring and the Probabilistic Method*, Algorithms and Combinatorics, vol. 23. Springer: Berlin.
- Neal RM. 2003. Slice sampling. *Annals of statistics* **31**(3):705–741.
- Ng F, Conway H. 2004. Fast-flow signature in the stagnated Kamb Ice Stream, West Antarctica. *Geology* **32**(6):481–484.
- Pattyn Frank, Derauw Dominique. 2002. Ice-dynamic conditions of Shirase Glacier, Antarctica, inferred from ERS SAR interferometry. *Journal of Glaciology* **48**(163):559–565.
- Rasmussen CE, Williams CKI. 2006. *Gaussian Processes for Machine Learning*. MIT Press: Cambridge, MA.
- Richardson C, Aarholt E, Hamran SE, Holmlund Per, Isaksson Elisabeth. 1997. Spatial distribution of snow in Western Dronning Maud Land, East Antarctica, mapped by a ground-based snow radar. *Journal of Geophysical Research: Solid Earth* **102**(B9):20343–20353.
- Rignot E, Bamber JL, van den Broeke MR, Davis C, Li Y, van de Berg WJ, van Meijgaard E. 2008. Recent Antarctic ice mass loss from radar interferometry and regional climate modelling. *Nature Geoscience* **1**(2):106–110.
- Rignot E, Mouginot J, Scheuchl B. 2011a. Ice flow of the Antarctic ice sheet. *Science* **333**(6048):1427–1430.

- Rignot E. 2006. Changes in ice dynamics and mass balance of the Antarctic ice sheet. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* **364**(1844):1637–1655.
- Rignot E, Velicogna I, Van den Broeke MR, Monaghan A, Lenaerts JTM. 2011b. Acceleration of the contribution of the Greenland and Antarctic ice sheets to sea level rise. *Geophysical Research Letters* **38**(5):L05503.
- Riva REM, Gunter BC, Urban TJ, Vermeersen BLA, Lindenberg RC, Helsen MM, Bamber JL, van de Wal RSW, van den Broeke MR, Schutz BE. 2009. Glacial isostatic adjustment over Antarctica from combined ICESat and GRACE satellite data. *Earth and Planetary Science Letters* **288**:516–523.
- Rue H, Held L. 2005. *Gaussian Markov Random Fields: Theory and Applications*. Chapman and Hall/CRC Press: Boca Raton, FL.
- Rue H, Martino S. 2009. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society B* **71**:319–392.
- Scott JBT, Gudmundsson GH, Smith AM, Bingham RG, Pritchard HD, Vaughan DG. 2009. Increased rate of acceleration on Pine Island Glacier strongly coupled to changes in gravitational driving stress. *The Cryosphere* **3**:125–131.
- Shepherd A, Ivins ER, Geruo A, Barletta VR, Bentley MJ, Bettadpur S, Briggs KH, Bromwich DH, Forsberg R, Galin N, Horwath M, Jacobs S, Joughin I, King MA, Lenaerts JTM, Li J, Ligtenberg SRM, Luckman A, Luthcke SB, McMillan M, Meister R, Milne G, Mouginot J, Muir A, Nicolas JP, Paden J, Payne AJ, Pritchard H, Rignot E, Rott H, Srensen LS, Scambos TA, Scheuchl B, Schrama EJO, Smith B, Sundal AV, van Angelen JH, van de Berg WJ, van den Broeke MR, Vaughan DG, Velicogna I, Wahr J, Whitehouse PL, Wingham DJ, Yi D, Young D, Zwally HJ. 2012. A reconciled estimate of ice-sheet mass balance. *Science* **338**(6111):1183–1189.
- Storvik G, Frigessi A, Hirst D. 2002. Stationary space-time Gaussian fields and their time autoregressive representation. *Statistical Modelling* **2**:139–161.
- Stroud JR, Müller P, Sanso B. 2001. Dynamic models for spatiotemporal data. *Journal of the Royal Statistical Society B* **63**:673–689.
- van Wessem JM, Reijmer CH, Morlighem M, Mouginot J, Rignot E, Medley B, Joughin I, Wouters B, Depoorter MA, Bamber JL, Lenaerts JTM, van de Berg WJ, van den Broeke MR, van Meijgaard E. 2014. Improved representation of East Antarctic surface mass balance in a regional atmospheric climate model. *Journal of Glaciology* **60**(222):761–770.
- Vaughan DG, Adalgeirsdóttir G, Agosta C, Bamber JL, Barletta V, Durand G, Edwards TL, Fettweis X, Gagliardini O, Hellmer HH, Holland PB, Howard T, Payne AJ, Ridley J, Smeets P, Spada G, Timmermann R, Trouvilliez A, van den Broeke M. 2013. *From Ice To High Seas: Sea-level Rise and European Coastlines*, Ford EAK (ed.): The ice2sea Consortium, Cambridge, United Kingdom.
- Velicogna I, Wahr J. 2013. Time-variable gravity observations of ice sheet mass balance: precision and limitations of the GRACE satellite data. *Geophysical Research Letters* **12**:3055–3063.
- Wikle CK, Berliner LM. 2007. A Bayesian tutorial for data assimilation. *Physica D: Nonlinear Phenomena* **230**(1–2):1–16.
- Wikle CK, Milliff RF, Nychka D, Berliner LM. 2001. Spatiotemporal hierarchical Bayesian modeling tropical ocean surface winds. *Journal of the American Statistical Association* **96**(454):382–397.
- Zammit-Mangion A, Rougier JC, Bamber JL, Schoen NW. 2014. Resolving the Antarctic contribution to sea-level rise: a hierarchical modelling framework. *Environmetrics* **25**:245–264.
- Zammit-Mangion A, Dewar M, Kadirkamanathan V, Sanguinetti G. 2012. Point process modelling of the Afghan War Diary. *Proceedings of the National Academy of Sciences* **109**(31):12414–12419.

APPENDIX A. PARAMETER ESTIMATION FROM NUMERICAL MODELS

A.1. Surface process parameters

For the surface layer, we estimated the Matérn and auto-regressive parameters from RACMO. In Zammit-Mangion *et al.* (2014), we provided estimates in the *spatial* setting by gridding the spatial domain and fitting homogeneous Matérn models to each block. Here, we adopt the same approach but instead fit a Gaussian separable *spatio-temporal* model to each block, with covariance function $k(u, \Delta) = k_1(u)k_2(\Delta)$. We maximise the likelihood using a grid-based method for each block (for $\nu = \{1, 2\}$) with 100 locations selected at random in each block. This provided estimates for a_S , κ_S , σ_S^2 and ν_S . For all blocks, $\hat{\nu} = 1$ was estimated.

The SPDE approach applied in Section 2.4 allows for smoothly varying parameters, whereby the local interpretation of the second-order properties are preserved. Therefore, in order to avoid discontinuities at the boundary, we subsequently obtained spatial maps for $a_S(s)$, $\kappa_S(s)$ and $\sigma_S^2(s)$ by interpolating the points using Gaussian radial basis functions with standard deviations of 600 km. The blocking strategy and the resulting maps for all spatially distributed parameters are shown in Figure A.1.

As seen in Figure A.1, the temporal scales vary considerably. We observe an overall positive dependence in the lower right (in the plotted sense) and upper parts of the continent while an overall negative temporal dependence in the central band. The spatial scales are equally irregular and conform with those observed in Zammit-Mangion *et al.* (2014). Scales range from around 80 km at the peninsula (top left

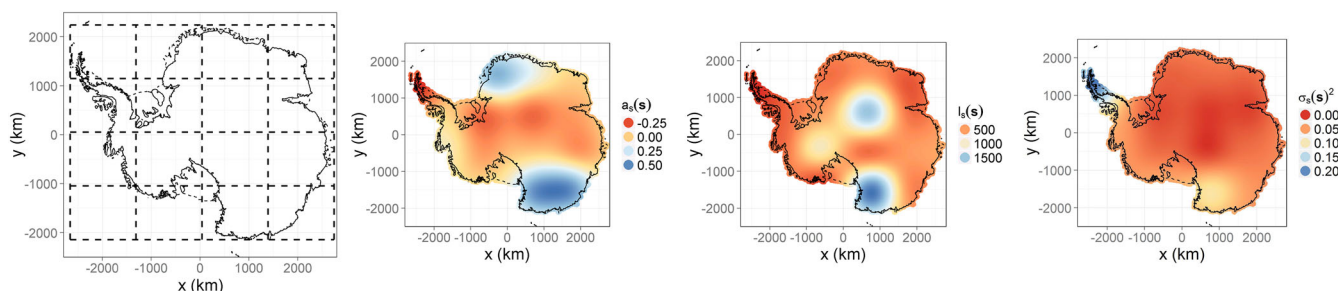


Figure A.1. Heterogeneity in surface process obtained by fitting local spatio-temporal separable models and interpolating using Gaussian radial basis functions. From left to right: (i) Gridding up of the domain in blocks, the data in which are assumed to be independent. Separate models are fitted within each of these blocks individually; (ii) propagation coefficient $a_S(s)$; (iii) spatial range $l_S(s) = \sqrt{(8\nu_S)/\kappa_S(s)}$ where $\nu = 1$; and (iv) marginal variance $\sigma_S^2(s)$

corner) to over 1500 km in the centre of the continent where precipitation is at a minimum. The marginal standard deviation further shows the spatially varying nature of height change due to surface processes and is as low as 2 cm year⁻¹ in the centre of the continent, 20 cm year⁻¹ on the coast and up to 40 cm year⁻¹ at the wetter peninsula.

A.2. Glacio-isostatic adjustment process parameters

The GIA model we use to obtain \mathbf{x}_M is that of Ivins *et al.* (2013), IJ05R2. We carry out a simple grid search on the likelihood function for $\nu = 1$ and $\nu = 2$, which is maximised for $\hat{\theta}_{GIA} = (0.0024, 7.3 \times 10^{-7}, 2)$. $\hat{\kappa}_{GIA} = 0.0024$ corresponds to a range of approximately 1700 km, lower than that reported in Zammit-Mangion *et al.* (2014) (3000 km), which considered an earlier version of the model (Ivins and James, 2005). $\hat{\sigma}_{GIA} = 0.85$ mm is also lower than the previous estimate of 4.2 mm and reflects the lower GIA rates in the revised model. Importantly though, the overall expected magnitude (mm) and length scales (>1000 km) are of the same order. This implies that temporally invariant, low power, spatially low frequency components of the data will be attributed predominantly to the GIA process.

A.3. Surface-firm interaction parameters.

Because we are assuming a coregionalisation model, an estimate for \mathbf{R} is obtained by finding the empirical covariance matrix between the surface and firm processes from RACMO and a firm densification model (Ligtenberg *et al.*, 2011) on a common grid. Here, we use the same estimate as that employed in Zammit-Mangion *et al.* (2014) given by

$$\hat{\mathbf{R}} = \begin{bmatrix} 1 & -0.41 \\ -0.41 & 0.24 \end{bmatrix} \quad (\text{A.1})$$

APPENDIX B. FINE-SCALE VARIATION HYPER-PARAMETERS

For a GMRF with fine-scale variation, the likelihood of θ_δ can be found analytically by integrating out \mathbf{x} and β_I . Rewriting the observation layer as $\mathbf{z} = \mathbf{C}'\mathbf{x}' + \delta + \mathbf{e}$, the log-likelihood is given by

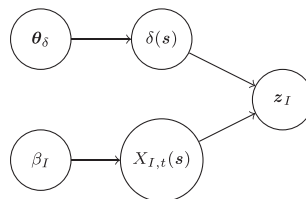
$$\ln p(\mathbf{z}|\theta_\delta) = c_1 + \frac{1}{2} \ln |\mathbf{Q}_{e\delta}| - \frac{1}{2} |\mathbf{Q}_{x'}| - \frac{1}{2} \mathbf{z}^T \mathbf{Q}_{e\delta} \mathbf{z} + \frac{1}{2} \mathbf{z}^T \mathbf{Q}_{e\delta} \mathbf{C}' \mathbf{Q}_{x'} \mathbf{C}'^T \mathbf{Q}_{e\delta} \mathbf{z} \quad (\text{B.1})$$

where $\mathbf{Q}_\delta = \text{diag}(\exp(-\eta_1 - b\eta_2))$, b is the roughness $b(s)$ evaluated at the observation locations, $\mathbf{Q}_{e\delta} = [\mathbf{Q}_e^{-1} + \mathbf{Q}_\delta^{-1}]^{-1}$, which is also diagonal,

$$\mathbf{Q}_{x'} = \mathbf{C}'^T \mathbf{Q}_{e\delta} \mathbf{C}' + \mathbf{Q} \quad (\text{B.2})$$

and c_1 is a constant independent of θ_δ .

Evaluating (B.1) for fixed θ and γ is very costly as the large footprint of GRACE serves to correlate all state variables *a posteriori*, resulting in dense matrices. For example, the Cholesky factor of $\mathbf{Q}_{x'}$, following fill-in reduction permutations, is over 5 GB when including GRACE observations, when compared to a few hundred Mb when using the point-referenced observations. However, we are solely interested in (B.1) to explore the parameter space and elicit a prior distribution over θ_δ . For this purpose, it is sufficient to consider the following simplified model:



That is, we consider only what ICESat (which has more coverage than EnviSat over rough terrain) is reading, assuming that all height change observed is due to ice dynamics or subshelf melting. We expect that the likelihood surface obtained in this way reflects an overestimation of θ_δ , because the other spatio-temporal processes (omitted in this simplified model) do explain some of the apparent variation. Using a simplified model allows us to explore the likelihood surface in considerable detail prior to applying a full MCMC scheme. The surface provides insights into how to transform the parameters for efficient sampling.

A plot of $\ln p(\mathbf{z}_I|\theta_\delta)$ against gridded values for η_1, η_2 is shown in Figure B.1 (left). The two parameters are, as expected, highly correlated because a large portion of the continent's surface is flat: the difference between a topographic-induced variation and a baseline variation is not pronounced. The surface plot, however, provides substantive evidence that the topographic-induced error is indeed the dominant component. We configured our prior based on this surface: we set $\mu_{\eta_1} = -12$ and $\mu_{\eta_2} = 2$, while for our uncertainty, because of use of the simplified model, we set $\sigma_{\eta_1}^2 = \sigma_{\eta_2}^2 = 1$.

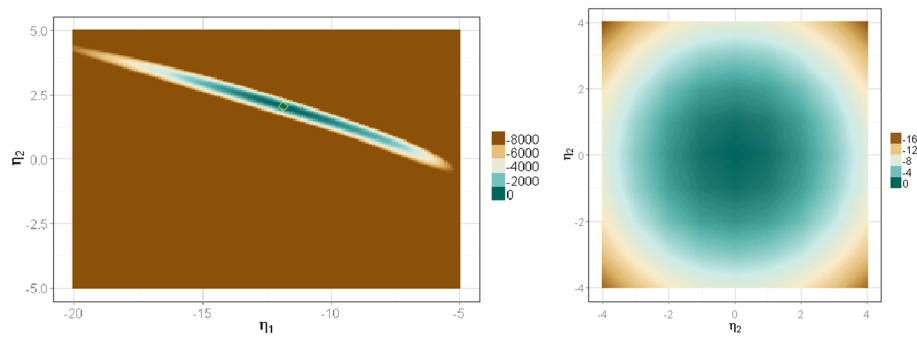


Figure B.1. Unnormalised log-likelihood of fine-scale variation parameters θ_δ shifted so that the maximum is zero. **Left:** Pretransformation, with the diamond denoting the maximum likelihood. **Right:** Post-transformation

We next transform the fine-scale parameters. Using the gridded maximum as a starting value, we carry out a gradient-free optimisation to find the maximum likelihood and then compute the Hessian numerically (using the R package `numDeriv`). We use the Hessian and the mode to transform the variables into (η'_1, η'_2) such that (η'_1, η'_2) are approximately zero-centred uncorrelated Gaussian random variables. A plot of the likelihood surface for (η'_1, η'_2) is shown in Figure B.1 (right). In the MCMC scheme discussed, we sample directly from (η'_1, η'_2) .